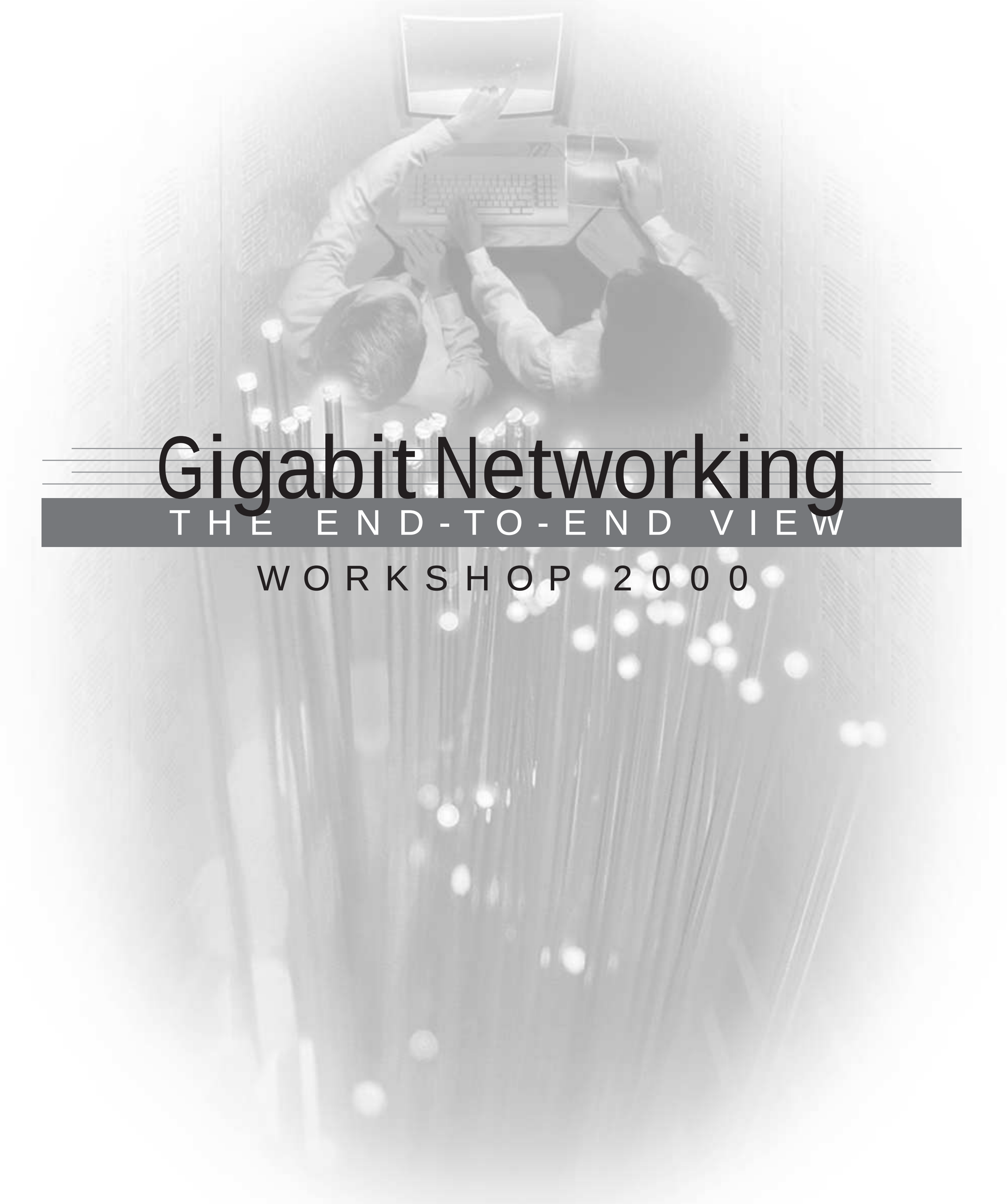
NASA AMES RESEARCH CENTER



Gigabit Networking

THE END-TO-END VIEW

WORKSHOP 2000

August 14 – 16, 2000

Welcome

On behalf of the NASA Research and Education Network (NREN), I welcome you to the San Francisco Bay area and to NASA Ames Research Center. Workshop 2000, “Gigabit Networking, the End-to-End View,” the fifth NREN Workshop, is a direct result of collaboration with other NGI agencies and research partners. The Gigabit Networking Organizing Committee has worked diligently to produce a venue for sharing gigabit networking status: notable accomplishments, interesting demonstrations, and current challenges. I thank each of you for taking advantage of this opportunity to shape the future of gigabit networking.

NREN and NASA are committed to advancing technology that not only promotes NASA’s mission in space but also benefits all of humankind. We know that building a team with other Federal agencies, academia and industry is the most expeditious way to meet the NGI goals.

We hope that this workshop experience is stimulating for you, the organizations represented and the NGI program. Again, thank you for participating and supporting this significant event.

Kevin L. Jones
NREN Engineering Group Lead

Table of Contents

Purpose and Rationale

The "Gigabit Networking: The End-to-end View" (GN) Workshop is devoted to examining the true status and realistic expectations for end-to-end gigabit-per-second ("gigabit") networking: infrastructure, applications, measurement, and the integration elements (e.g., middleware, tools, management) needed to make it all work together end-to-end. The workshop is hosted by the NASA Research and Education Network (NREN) at the request of the Large Scale Networking (LSN) Coordination Group.

The purpose of the workshop is to understand and achieve multiagency views, coordination, consensus, issues and recommendations, concerning where the LSN agencies and partners are currently and where they should be.

The specific objective of the workshop is to develop and publish "roadmaps" in selected focus areas. These roadmaps will describe paths to achieving at least ten true end-to-end gigabit networking application demonstrations by September 2002, which marks the end of the Next Generation Internet (NGI) program. Each roadmap will identify requirements, current status and issues, as well as describe what needs to be developed and deployed, according to the milestones and deliverable results that can be expected for public demonstration and measurement of applications performance by the end of the period.

The first half-day of the workshop, Monday, August 14th, is devoted to demonstrations of gigabit networking applications. Each demo team will also present a look "under the hood," identifying their particular implementation issues, approaches and remaining challenges.

The second day of the workshop, Tuesday, August 15th, is a plenary session intended to present the breadth of the salient gigabit networking issues.

On the third day, the draft workshop roadmaps will be developed in eight breakout sessions. In the morning, four Technology breakouts will identify the developments, challenges, and planned availability for the main technology building blocks of gigabit networking over the next two years. In the afternoon, four Application breakouts will focus on the main classes of gigabit networking applications with the aim of identifying their requirements, unique characteristics and demonstrations planned or hoped for within the next two years.

Your active help is needed to make this workshop a success. The first day demonstrations and presentations will be accessible only in the main conference room. The second day plenary session is being multicast to Internet2 and other partners. The third day results of the breakout sessions, i.e., the roadmaps which will constitute the principal output of the workshop, including analysis and recommendations for achieving the gigabit networking goals of the NGI program, will be published on the web in the public domain.

Thanks for everyone's efforts, both the organizers and the developers, in preparing for and participating in this very important workshop.

Richard desJardins, NASA/NREN

Workshop Arrangements

Information:

Staff are available at the Registration Desk to answer questions.

Messages:

The Registration Desk will accept messages for workshop participants at 1-650-604-2206. Messages will be delivered to the recipient.

There are phones throughout the facility for your use. Dial 7 to get an outside line.

Communications and message center:

Ethernet ports are available in the Fireside room adjacent to the Registration Desk to receive email.

Video taping:

The Plenary session presentations will be video-taped and multicast for the NGI public record.

Restrooms:

Restrooms are available on the north side of the Ballroom as well as off the front lobby (Main Entrance).

Agenda

Monday Afternoon, August 14

1200 – 1330 **Registration/Lunch**

1330 – 1340 **Welcome and Logistics**

1340 – 1540 **Demonstration Presentations**

1540 – 1600 **Break**

1600 – 1730 **Demonstration Presentations, cont.**

1730 – 1930 **Networking Reception**

Tuesday, August 15

0730 – 0830 **Continental Breakfast**

0830 – 0900 **Welcome and Logistics**

– Robert Rosen, Ames Center Associate Director for Aerospace: Welcome to Ames

– William Van Dalsem, NASA HPCC: The Importance of Gigabit Networking to NASA Enterprises

– Ken Freeman, NASA NREN: Demonstration Summary and Keynote Introduction

0900 – 1000 **Keynote: Raj Jain,** Ohio State University/Nayna Networks

1000 – 1030 **Break**

1030 – 1200 **Setting the Stage Panel**

– John Wroclawski, Massachusetts Institute of Technology

– Kay Howell, National Coordination Office for Computing, Information and Communications

– Richard Carlson, Department of Energy

– Phillip Dykstra, WearOnEarth Communications, Inc.

– Ian Foster, Argonne National Laboratory

– Wesley Kaplow, Qwest Government Systems

1200 – 1330 **Lunch**

1330 – 1500 **Developing the Infrastructure**

– Jim Gimlett, Network Elements: Platforms

– Leonid Kazovsky, Stanford University: Optical Access

– Nick McKeown, Stanford University: Fast Routing and Switching

– Basil Irwin, University Corporation for Atmospheric Research, & Matt Mathis, National Laboratory for Applied Network Research: Web100

1500 – 1530 **Break**

1530 – 1700 **Tools for Gigabit Networking**

MEASUREMENT

– Joe Evans, University of Kansas

GIGABIT END-TO-END ISSUES PANEL

– Steve Corbato, Internet2: Overview

– Michael O'Connor, Brookhaven National Laboratory: Campus

– Cas D'Angelo, Southern Crossroads Gigapop: Gigapops

– Jerry Sobieski, Mid Atlantic Crossroads: Peering

– William Lennon, Lawrence Livermore National Laboratory: Wide Area

1700 – 1730 **Breakout Session Instructions**

Agenda cont.

Wednesday, August 16

0730 – 0830 **Breakfast**

0830 – 1230 **Technology Breakout Groups - Session 1**

- Gigabit Testbeds
- Platforms
- Measurement
- Middleware/Integration

1230 – 1330 **Lunch**

1330 – 1530 **Application Breakout Groups – Session 2**

- Teleseminars/Telemeetings
- Models in Real Time
- Huge Databases
- Remote Instrumentation

1530 – 1600 **Break**

1600 – 1700 **Breakout groups reporting and workshop wrap-up**

Abstract

Gigabit Networking Workshop,
August 14-16, 2000
NASA/NREN NGI Workshop 2000:

"Gigabit Networking: The End-to-end View"

NASA Ames Research Center,
Conference Center (Building 3)
Moffett Field, California 94035
Contact Kevin Jones: *kjones@arc.nasa.gov*

– Monday, August 14 afternoon/evening: Demonstrations/reception

– Tuesday-Wednesday, August 15 morning to August 16 evening: Workshop

– Thursday, August 17: Splinter meetings and report writing

HOSTED BY NASA RESEARCH AND EDUCATION NETWORK (NREN), ON BEHALF OF THE LARGE SCALE NETWORKING WG AND ITS TEAMS

Participation is by invitation only.

The NREN Next Generation Internet (NGI) Workshop 2000 is devoted to examining the true status and realistic expectations for end-to-end gigabit-per-second (Gbps) networking: connectivity, platforms, applications, measurement, and the integration aspects (e.g., middleware, tools, management) needed to make it all work together end-to-end. As in past NGI workshops hosted by NREN, participation in the workshop is by invitation only, divided among government and industry NGI technologists, university researchers, agency NGI representatives, and policymakers and advisors including representatives of the President's Information Technology Advisory Committee (PITAC).

PURPOSE AND OBJECTIVES

The purpose of the Gigabit Networking Workshop is to understand and achieve multiagency views, coordination, consensus, issues, and recommendations, concerning where the NGI agencies and partners are currently, and where they are (or should be) going, in order to demonstrate true end-to-end Gbps networking by the end of the NGI program.

The specific objective of the workshop is to develop and provide "roadmaps" in each of the theme areas of the workshop. These roadmaps will describe paths to achieving at least ten true end-to-end Gbps networking application demonstrations by the end of the NGI program in September 2002. Each roadmap will identify requirements, current status and issues, as well as describe what needs to be developed and deployed, according to what milestones, and what deliverable results can be expected for public demonstration and measurement of applications performance by the end of the period.

TECHNICAL PROGRAM CONTENT AND PROCESS

Technical program content of the workshop is divided into five themes:

– *Connectivity Infrastructure:* How do you move gigabits-per-second end-to-end?

– *End-system Platforms:* What do you put at the ends of the connections? (e.g., computer architectures, computer hardware, OS, cards, chips)

– *Applications:* What application software runs in the platforms, and what for?

– *Measurement:* What and how do you measure for Gbps end-to-end performance?

– *Integration:* What else is needed to make all of this work? (e.g., security, QoS, multicast, middleware, management).

The aim of the technical program content is to cover every essential topic needed to achieve true Gbps networking end-to-end, from one application user to one or more other users or application service providers.

Abstract cont.

DEMONSTRATIONS

Demonstrations have been invited from the NGI agencies and partner communities. The goal is to have four to six good demos, each demonstrating 500 Mbps+ throughput. Full OC-12 bandwidth will be available to each demonstration by scheduling timeslots for each demo (or perhaps by QoS mechanisms, as a research issue). The aim is to demonstrate real applications that “fill the pipe,” achieving substantially Gbps rates end-to-end.

Some of the demo teams will also “open the hood” and show end-to-end how their team approached the engineering issues involved in their networking application demo, and how much they aim to improve their end-to-end performance in the next two years.

WORKSHOP MANAGEMENT

The NREN management team for the workshop is headed by Kevin Jones, and is divided into four areas: putting on the demonstrations; conducting the workshop technical program; writing the report (including “so what?” analyses of the demos, and “now what?” results of the workshop); and local arrangements and logistics.

The workshop reporting aim is to provide the first draft of the workshop report immediately following the workshop. NREN will provide a team of technical writers and NREN shadows organized to support the workshop leads in achieving this goal. The workshop report will include identification of participants, and copies of their presentations will be published on the NREN web site.

Organization of the workshop is carried out by an Organizing Committee chaired by NREN, and includes members from the NGI agencies, LSN teams, and Internet2.

Executive Summary

1 Workshop Purpose and Goals

The "Gigabit Networking: The End-to-End View" (GN) Workshop held at NASA Ames Research Center (ARC) on August 14-16, 2000, was devoted to examining the status of end-to-end gigabit-per-second networking and determining what advances can realistically be achieved within the next two years. Specific areas addressed at the workshop include infrastructure, applications, measurement, and the integration elements (e.g., middleware, tools, management) needed to make it all work together end-to-end. The workshop was hosted by the NASA Research and Education Network (NREN) at the request of the Large Scale Networking Coordination Group (LSN).

The specific objective of the workshop was to develop and publish "roadmaps" in selected focus areas, both technology areas and application areas. These roadmaps describe paths to achieving ten or more true end-to-end gigabit networking application demonstrations by September 2002, which marks the end of the Next Generation Internet (NGI) Federal Program. Each roadmap contains information regarding current status of the technology or application area as well as a discussion of issues, challenges and recommendations.

During the first half-day of the workshop, Monday, August 14, state-of-the art gigabit networking applications were demonstrated and described. Each demo team exposed its demonstration "under the hood," identifying particular implementation issues, approaches and remaining challenges.

The second day of the workshop, Tuesday, August 15, was a plenary session that presented the breadth of the salient gigabit networking issues.

On the third day, the draft workshop roadmaps were developed in eight breakout sessions. In the morning, four Technology breakouts identified the current status, ongoing developments, challenges and future plans regarding the main technology building blocks of gigabit networking over the next two years. In the afternoon, four Application breakouts focused on major classes of gigabit networking applications in order to characterize their unique requirements, highlight existing challenges, and identify demonstrations planned or hoped for within the next two years.

The first day demonstrations and "under the hood" presentations were accessible only in the main conference room. The second day plenary session was multicast to Internet2 and other partners; the plenary presentations are published on the web at *www.nren.nasa.gov/gn_agend.html*. The results of the third day breakout sessions, i.e., the roadmaps which constitute the principal output of the workshop, including analysis and recommendations for achieving the gigabit networking goals of the NGI program, are published in the Workshop Report at *www.nren.nasa.gov/gn_report.html*.

The active participation of all attendees helped to make this workshop a success. Thanks for everyone's contributions, especially the organizers, developers and presenters, for making this workshop a success. The remainder of this Executive Summary provides a brief description of the workshop outcomes. For the entire report and presentations, please visit the web sites given above.

2 Demonstrations

Seven live network demonstrations representing five different agencies and Internet2 were presented on the first day of the workshop. These were all shown on the 20’ x 24’ screen of the main auditorium so that everyone could see the demonstrations and hear the “under the hood” explanations.

Due to continuing deployment delays, the National Transparent Optical Network (NTON) was able to provide only OC-12 (622 Mbps) service instead of the desired OC-48 (2.5 Gbps) service to NASA ARC. Accordingly, minimum end-to-end performance of 200 Mbps was established as the baseline for selection of the workshop demonstrations. At least one demonstration, the Project Data Space, utilized over 250 Mbps.

A summary of each demonstration follows. For more details, see *www.nren.nasa.gov/gn_appli.html.*

2.1 DIGITAL SKY VIRTUAL OBSERVATORY

(NASA JET PROPULSION LABORATORY)

In the Digital Sky Virtual Observatory, three large data streams from geographically distributed sites were routed over high speed networks, combined on the fly and projected for viewing at ARC using Virtual Observatory software. The sources of the data streams were the Infrared Astronomy Satellite (IRAS) from Goddard Space Flight Center (GSFC), a Digitized Palomar Observatory Sky Survey/Two Micron Sky Survey (DPOSS/2MASS) composite from the Jet Propulsion Laboratory (JPL), and the 2MASS center of the galaxy mosaic from CalTech. Also shown was the 200-gigabyte U.S. mosaic served from JPL. The demo required bringing together server- and client-side applications, networks spanning the country, and supercomputers at ARC, JPL, CalTech, and GSFC.

2.2 PROJECT DATA SPACE

(UNIVERSITY OF ILLINOIS/NATIONAL SCIENCE FOUNDATION)

Project Data Space linked 12 sites across five continents to demonstrate a new infrastructure to handle 1) remote data access, analysis, and mining, and 2) distributed data analysis and mining. Led by researchers at the University of Illinois at Chicago, the team demonstrated a variety of tools using its new Data Space Transfer Protocol (DSTP) to publish, access, analyze, correlate and manipulate remote and distributed data.

2.3 VISIBLE HUMAN

(UNIVERSITY OF MICHIGAN/NATIONAL INSTITUTES OF HEALTH)

The Visible Human (VH) Project® creates complete, anatomically detailed, 3-D representations of the normal male and female human bodies and enables interactive biomedical image segmentation, labeling, classification, and indexing of large images. In the VH demonstration, images were rendered on an SGI at the University of Michigan and transmitted across the Internet2/Abilene backbone in near real time to NASA ARC. The VH Project demonstrated the use of gigabit Ethernet and a high-speed network backbone to allow a user to navigate in near real time through selected views from anatomic data segments.

2.4 COMBUSTION CORRIDOR—VISAPULT

(LAWRENCE BERKELEY NATIONAL LABORATORY/DEPARTMENT OF ENERGY)

The Combustion Corridor—Visapult project from DOE demonstrated remote and distributed visualization of tera-scale scientific data using a combination of parallel network storage caches and parallel and cooperative rendering. Visapult couples resources that may be separated by

2 Demonstrations cont.

distance, and uses a novel architecture for parallelizing data communication and for parallel and cooperative rendering. The core of the application framework is the network.

2.5 INTERNET2/LAND SPEED RECORD

(DARPA, PRESENTED BY INFORMATION SCIENCES INSTITUTE)

The Internet2 Land Speed Record demonstration reviewed current gigabit rate applications within SuperNet and demonstrated cross-OS high bandwidth data transfers and the impact of maximum transmission unit (MTU) sizes on throughput. Even with very high-end workstations, it is still necessary to use large MTU sizes to push beyond 700 Mbps. Given the widespread use of 4470 MTU sizes on NGI core OC48 links, it clearly will be difficult to move beyond 750 Mbps in the near term.

2.6 UNIVERSITY OF WASHINGTON HIGH DEFINITION TV

(DARPA, NSF, SONY)

The University of Washington Internet HDTV Project demonstrated software capable of streaming studio-quality HDTV at speeds in excess of 200 Mbps. The project explores the intersection of network, video, and server technologies where near real-time distribution of extremely high-quality images is required. A core design principle of the Internet HDTV project is reliance upon standards from the video and networking communities.

2.7 VIRTUAL MECHANOSYNTHESIS

(NASA AMES RESEARCH CENTER)

Virtual Mechanosynthesis (VMS) is a software application that allows collaborative, distributed computational-steering of a live parallelized molecular dynamic simulation. To provide effective visualization and steering of a remote simulation to multiple remote users in real time, consistent low latency and high bandwidth are required.

3 Technology Road Maps

The four technology areas that were the focus of breakout sessions were selected based on their significance for achieving the NGI end-to-end gigabit networking goals within the next two years. Each area represents a critical link in the end-to-end chain. Today people engineer specific solutions to enable one-time demonstrations. Without substantial technical advances in each of these areas, true end-to-end gigabit networking will not be achieved in a form that allows persistence and continuing evolution following the end of the NGI program in September 2002. The results of these technology breakout sessions are summarized below and given in full at *www.nren.nasa.gov/gn_break.html.*

3.1 GIGABIT TESTBEDS

The objective of the Gigabit Testbed breakout session was to establish a roadmap describing how to proceed with gigabit testbeds through the end of the NGI program in September 2002, and to discuss how testbeds should evolve post-NGI. Current gigabit networking access includes the national education intranet Internet2/Abilene and the DARPA SuperNet. The next step for many campuses would be the upgrade of their WAN access links from OC3 and OC12, yet the additional costs are prohibitive for most campuses. Gigabit networking to the desktop will require providing assistance to help campuses upgrade their LAN architectures and their WAN access links. In wide area networking, a major concern is the evolution and upgrading of NGIXs to a minimum of OC12 ATM interconnects. Many issues, such as testbed persistence and expansion, are directly related to funding issues best addressed by the funding agencies. It is recommended that the agencies and Internet2 establish a "gigabit strike force" to identify and publicize best practices in these areas, and for LSN teams to hold a workshop in early 2001 to ensure that progress towards gigabit goals remains on track.

3.2 PLATFORMS

"Platform" was defined as ending where the network begins—at the network interface card, which is within both platform and network—and as having two major constituents, hardware and software. Of these two, hardware is in reasonably good if uneven shape, but software needs work. End-to-end gigabit networking, while possible today, requires skill to perform significant amounts of manual tweaking of the software, primarily in the OS and network stack. Careful attention to preventing packet loss is necessary to achieve gigabit rates even with carefully tuned platforms. The issue of whether to require jumbo frames remains unresolved. It is recommended that agencies and universities perform extensive experimentation and documentation of lessons learned in this area, and that open source autotuning software (e.g., Web 100+), "cookbooks," and other easy-to-use and freely distributed documentation be aggressively developed and widely publicized.

3.3 MEASUREMENT

Many aspects of network measurement are common to all networks regardless of speed. Examples of measurement challenges unique to gigabit networking include difficulty in collecting and analyzing high speed data and the potential performance degradation caused by small error rates. The breakout session addressed both aspects of network measurement. Some of the conclusions reached in the breakout session are presented below.

3 Technology Road Maps cont.

Existing measurement tools can serve for rudimentary assessment of end-to-end performance, but capabilities are missing for network segmentation and application-specific characterization of problems. The speed of network infrastructure growth and introduction of new transport technologies means that no single tool or small set of fixed tools will provide adequate measurements. Very small error rates can have a significant impact on the performance of high bandwidth long-distance networks; the measurement tools must be able to detect and report what used to be considered insignificant. An integrated network measurement service framework that is adaptable at the hardware and software level is essential to provide scaling and to accommodate ever-changing transports.

3.4 MIDDLEWARE/INTEGRATION

“Middleware” represents a broad and diverse set of tools to facilitate the use of basic network capabilities by applications and users. Gigabit networking applications should aggressively experiment with the use of “grid” middleware (e.g., Globus) and the core middleware services defined by the Internet2 Middleware Initiative. These middleware elements will provide important services such as identification, authorization, authentication, and directories.

“Integration” is a catchall term that includes the miscellaneous elements essential to the delivery of gigabit to the desktop that are not covered elsewhere. There are many challenges to integration: Individual developers tend to engineer their own solutions to make their application work; gigabit networking to the desktop tends to work beyond the documented and tested limits of hardware and software; undocumented pathological behaviors tend to emerge, causing problems that are difficult to diagnose; performance tuning is critical; and interoperability is key. Deployment of community-agreed middleware and services such as identified above would be a giant step towards much easier integration of end-to-end gigabit networking applications. Other steps would be the creation of a common trouble ticket system across networking domains, and development of tools for mapping identifiers and attributes across domains.

4 Application Road Maps

The four application areas that were the focus of breakout sessions were selected based on their unique requirements and their importance to agencies, university researchers and the public. Taken together, we believe that these application areas span the important application space for gigabit networking in the next two years. To achieve true gigabit networking, each area must be demonstrated in a form that will allow persistence and continuing evolution following the end of the NGI program in September 2002. The results of these application breakout sessions are summarized below and given in full at *www.nren.nasa.gov/gn_break.html.*

4.1 TELESEMINARS/ TELEMEETINGS

Teleseminars are remote interactive lectures and presentations, requiring one-to-many multicasting, very high video resolution and excellent audio fidelity primarily in one direction. Telemeetings are remote interactive group discussions and presentations, requiring potentially all-to-all multicasting in addition to raising difficult session control and feedback issues. Both applications are significant traffic generators and raise substantial service and performance issues such as bandwidth, traffic engineering, errors, latency, multicast, control, and measurement.

“Extreme conferencing” is an informal term for the maximum conferencing capability achievable; the sole example is the DARPA Virtual Amphitheater, which aims to bring hundreds of moderate resolution participant images plus several high resolution central stage or panel images into a real-time virtual reality amphitheater. Studio-quality high definition television (HDTV)-based conferencing is needed to support full quality teleseminars such as for surgery; University of Washington, DOD, NASA, NIH and NOAA all have excellent applications in this area. Room-based conferencing requiring at most one unskilled operator is available now but needs to be upgraded substantially in audio and video quality and resolution; the Internet2 Access Grid and cooperative agency/industry HDTV-based videoconferencing prototyping may produce gigabit networking examples. “Cookbook” desktop conferencing of spontaneous multi-participant sessions and requiring no operators is not currently a potential true gigabit networking application; examples are VIDE and VRVS.

4.2 MODELS IN REAL TIME

Models in real time refers to the human steering or visualizing of one or more remote supercomputing models in highly interactive modes and time scales. The interesting gigabit networking problems in this area arise because supercomputing modeling is currently primarily a batch operation, with carefully prepared inputs, compute times frequently of hours or even days, and then relatively long cycles of human visualization and analysis of the model outputs.

Two examples of models in real time were shown at the conference. DOE demonstrated the Combustion Corridor, which models internal combustion with a wide variety of modes and parameters steered and visualized by remote researchers in real time and near real time. NASA demonstrated Virtual MechanoSynthesis (VMS), which allows researchers to interactively manipulate nanoengineering models of carbon and other molecular level structures. Also, DOD distributed interactive battlefield simulations of tens of thousands to hundreds of thousands of elements should reach gigabit aggregate rates within the next two years. NOAA and DOE have very interesting potential applications in this area, but it appears that lack of targetable

4 Application Road Maps cont.

funds will prevent them from pursuing work in this area during the next two years.

4.3 HUGE DATABASES

The “Huge Databases” application area involves remote extractions of up to terabytes of data from distributed petabyte databases in human-scale time. Extraction and analysis of data from such databases requires massive processing resources (multiprocessing minisupercomputers, which may also be remote and distributed) as well as terabytes of dedicated local storage and a high resolution local graphics engine for local imaging or rendering of the data.

Earth science and “digital earth” databases from NASA, NOAA and DOD programs are excellent examples, particularly in the era of the terabytes per day of data generated by the Earth Observing System (EOS) and other space-based and in-situ earth observing sensors. Databases from DOE high energy physics experiments, as well as NIH and DOD digital libraries of anatomical and radiological images and models, also represent “huge” databases that are accessible in human scale time only with the support of gigabit networking services.

4.4 REMOTE INSTRUMENTATION

Remote instrumentation refers to the use of scientific instruments operated remotely over a network. Requirements for human interactive feedback focus on maintaining latency (i.e., end-to-end delay) at a nearly constant specified minimum, ranging from a few milliseconds for haptic (i.e., “touch and feel”) feedback for telesurgery, to a few tens of milliseconds for adjusting microscope specimens or focus settings in real time, to a few hundreds of milliseconds or a few seconds for command verification for telescope instruments. Remote instrumentation typically also requires stringent security and support for differentiated services (e.g., different network service guarantees for image streaming versus commanding).

Examples of remote instrumentation demonstrations requiring gigabit networking include NSF-supported remote telemicroscopy activities taking place at San Diego Supercomputer Center (SDSC); NIH- and DOD-supported remote telesurgery, including surgery training and planning using simulation; and remote astronomical telescopes (NSF and NASA), oceanographic robotic exploration operations (NOAA and DOD), and remote robotic telepresence in hazardous environments (DOE and EPA).

5 Next Steps

The consensus following the workshop presentations and discussions was that true end-to-end gigabit networking (defined as at least 500 Mbps end-to-end sustained performance delivered to the user) is not yet here, but it's close. The issue is not that the raw network physical layer technology is not capable of these rates, but that the end-to-end multilayer deployment and integration of applications together with networking services is currently still a hand-crafted and nearly always "stovepipe" activity.

The situation is improving quickly, however, and we credit Professor John Wroclawski of MIT with the observation that we're at the stage where we no longer have to be a *guru* to do gigabit networking, but we still have to *know a guru*. By the end of the NGI program, LSN would like to be at the point that gigabit networking is persistent, continually evolving, and is *becoming routine* at major Federal laboratories and research universities. This will mean (by definition, since the agencies and universities buy their network services and equipment from industry) that gigabit networking is "ready for prime time."

Specific next steps for the technology and application areas discussed at the workshop are given in the roadmaps in the full report. We believe that the goal of ten or more true end-to-end gigabit networking demonstrations (as defined by the above criteria) over nationwide distances is quite achievable by September 2002.

LSN should delegate to its *High Performance Network Applications Team (HPNAT)* the task of identifying hopefully as many as 20 gigabit networking applications that the agencies and Internet2 partners will commit to carrying out successfully, with the goal of persistence and continuing evolution afterwards, by the end of this period.

The LSN *Joint Engineering Team (JET)* would then be responsible for assuring the end-to-end testbed interoperability and persistent infrastructure to deploy and demonstrate these applications.

Research issues identified in the detailed technology session reports should be examined specifically by the LSN *Networking Research Team (NRT)*, with a view towards developing and recommending solutions to agencies for deployment in support of the selected applications.

Specific areas of focus for LSN and its teams should be the *measurements* required to prove the continuing achievement of the gigabit networking performance, and the *middleware* needed to be deployed in order to create a persistent, interoperable, end-to-end infrastructure across the networks of multiple agencies and Internet2.

LSN should review the progress of its agencies and Internet2 in achieving the above gigabit networking goals by having an *interim review meeting within a year* and preferably within nine months. This review meeting should cover:

- Identification of up to 20 specific gigabit net working applications, including demonstration milestones;
- Specific lead agency plans for achieving the milestones;
- Critical research issues and activities needed to achieve the milestones;
- Significant deployments of interoperable middleware needed;
- Two-year funding availability for each application; and
- Agency and Internet2 plans for knowledge repositories in the gigabit networking area, so that key areas that become routine will have web sites, "cookbooks," and lessons learned teleseminars made available in the public domain.

4 Demonstrations

Digital Sky Virtual Observatory

Project Data Space

Visible Human Images

Combustion Corridor—Visapult

Internet2/Land Speed Record

University of Washington HDTV

Virtual MechanoSynthesis

Digital Sky Virtual Observatory Demonstration Over NREN

JOSEPH JACOB, JET PROPULSION LABORATORY

In this demonstration, two large data streams from geographically distributed locations will be routed over high speed networks, combined on the fly, and viewed on a high resolution Powerwall display at Ames Research Center (ARC) using the Jet Propulsion Laboratory's (JPL) Digital Sky Virtual Observatory software, described below. The source of these data streams will be two large image mosaics constructed from Digital Sky datasets. The joined image will be a multi-spectral, multi-resolution representation of the sky. The visualization software will run as a client on an SGI machine at ARC. This client will connect over the networks to JPL image server software running on SGI machines at each of the remote sites. The sites that may serve the data are to be determined, but may include JPL and Goddard Space Flight Center (GSFC). The following is additional background on the Digital Sky Virtual Observatory project and the organizations participating in this demonstration.

One of the first steps towards creating a National Virtual Observatory is to allow astronomers and researchers an intuitive, comprehensive way to remotely navigate the immense datasets produced by various all sky surveys. The objective of the Digital Sky Virtual Observatory Project is to design and prototype a system capable of accessing and interrelating large, geographically distributed datasets from multiple sky surveys including images and the relation to catalog information.

The Digital Sky Virtual Observatory application developed at JPL allows one to view higher-resolution insets of particular regions of the sky overlaid on mosaics of the entire sky. This application is based on an enhanced multi-screen Electronic Light Table (ELT) technology developed by the JPL Advanced Laboratory for Parallel and High Performance Applications (ALPHA) team, which is based on a single screen development visualization software, ELT, from SGI. With this software, users are able to pan and zoom from the full sky into a portion of it where the full resolution, typically an arc second, can be seen. The all sky surveys consist of images and catalog data. The relation of the catalog information to the images and vice versa is critical to the usefulness of the tool. The catalog tool allows one to choose a celestial body from a catalog list and have the corresponding location marked on the mosaic, or to choose a location or region on the mosaic and have the corresponding catalog information highlighted in the catalog list.

JPL focuses on applying high performance computing and communication technology to visualize and analyze the Digital Sky datasets, both imagery and catalog data, at arbitrary resolutions. NREN works with the JPL ALPHA team to evaluate and implement remote capabilities suitable for this application. The ultimate goal of this partnership is to allow a user linked by high-speed networks to the appropriate high-end computational resources to view images created from data at remote data storage facilities.

http://alphabits.jpl.nasa.gov/DPAT/tasks.html#DSVOproj

Project Data Space

ROBERT GROSSMAN, UNIVERSITY OF ILLINOIS AT CHICAGO

Project Data Space will link 14 sites across 5 continents to demonstrate a new infrastructure to handle 1) remote data access, analysis, and mining, and 2) distributed data analysis and mining. Led by researchers at the University of Illinois at Chicago, the team will demonstrate a variety of tools using its new Data Space Transfer Protocol (DSTP) to publish, access, analyze, correlate and manipulate remote and distributed data. The team hopes that the DSTP infrastructure will provide the same ease of use for distributed data analysis and data mining that HTTP provided for viewing remote documents.

We will showcase DSTP Servers, DSTP Clients, and a variety of DSTP applications. The applications also use the Predictive Model Markup Language (PMML), an emerging standard for statistical models. For example, in the Sky Survey application, the DSTP Client downloads stellar object catalog data from a DSTP Server, creates a machine learning model based on PMML, and scores large amounts of data at high rates using high performance DSTP applications we have developed. Last year, we were able to move 250 Mbits/sec (~113 GB/hr) from our lab at the University of Illinois (Chicago) to the SC-99 showroom floor in Portland with no network tuning. We expect even higher rates at the NREN demonstration using a new release of our software.

The Network Storm application will demonstrate the flexibility of the DSTP Protocol. Locations will be set up on three continents to collect network traffic data. DSTP Servers are already installed at those locations and any person utilizing the DSTP protocol can download data and view the state of the network. Ultimately the group plans to build an infrastructure that will predict network storms and allow for improved network traffic management.

We will also demonstrate several other high performance DSTP applications.

IMPLEMENTATION APPROACH

The DataSpace Testbed consists of DataSpace Nodes, which are workstation clusters consisting of 2, 4, or 8 nodes, together with 50 GBs – 250 GBs of disk. These nodes are distributed in the US, Europe and Asia. We are utilizing a load-balancing switch, along with a GigE switch, to manage the work flow to these DataSpace servers.

Our approach is to use commodity equipment to achieve high performance data mining across a high performance WAN network (Internet2) without any network tuning. The four sites for this demonstration will each have a Terabyte Challenge node (3 Intel boxes with 200GB raid level 5 disk).

LESSONS LEARNED

Our approach has been to introduce a new protocol (the DataSpace Transfer Protocol) to enable next generation wide area, data intensive applications. We are currently exploring moving a variety of different data types using this protocol. We are experimenting adding additional capabilities to the protocol to be able to work effectively with different data types.

TCP/IP, as is, is not a high performance networking protocol. It is important to find new ways to achieve high performance without resorting to specialized equipment or network tuning. Our initial approach has been to write a library called PSockets that utilizes striping across multiple sockets to get around the latency issues of TCP. In testing, this has allowed for performance achieved over a WAN that is comparable to that which can be achieved over a LAN.

http://www.dataspaceweb.net

http://www.ncdm.uic.edu

Visible Human

BRIAN ATHEY, UNIVERSITY OF MICHIGAN

The Visible Human (VH) telemedicine demonstration is a first step in accessing large image database services over high-speed computer networks. A digital image library of volumetric data representing a complete, normal adult male and female cadaver (*The Visible Human Project*) currently resides at the National Library of Medicine (NLM) in Maryland. These thinly sliced images are of cryosections derived from computerized tomography and magnetic resonance. The demonstration implements a model that enables interactive biomedical image segmentation, labeling, classification and indexing to take place using large images.

The VH dataset is an information-rich dataset that does not exist in private sector datasets. Commercial subsets of the VH dataset are often compressed by lossy techniques, thus reducing information. Maintaining a centralized repository accessed over high-performance networks simplifies management of the database. Biomedical image libraries (in number and size) are sure to grow. Currently, licensees of the VH dataset number 1000+ worldwide. Due to the size and international importance of the dataset, multilingual labeling of the dataset is beginning to be carried. Therefore, various researchers will need interactive access to provide image segmentation and labeling. In the future, online access to an anatomical segmented human anatomy atlas will be a vital resource for biomedical researchers worldwide.

http://www.nlm.nih.gov/research/visible/visible_human.html

Combustion Corridor— Visapult: A Prototype Application and Framework for Remote and Distributed Visualization of Large Scientific Data

WESLEY BETHEL, LAWRENCE BERKELEY NATIONAL LABORATORY

As part of the DOE-sponsored Combustion Corridor research project, we have developed a prototype application and framework called Visapult that is used for remote and distributed visualization of large scientific data sets. As a framework, Visapult couples resources that may be separated by distance, such as network storage caches and parallel computing platforms. It uses a novel architecture for parallelizing data communication, and for parallel and cooperative rendering. The three primary components of the system are a viewer that may be run on an arbitrary Unix platform that supports OpenGL graphics, a back end that runs on a parallel computing platform, and a data source, such as a network storage cache. The motivation for this work stems from challenges resulting from ever-increasing data set sizes, and the researchers' need to view the results of their simulations.

At the core of the application framework is the network. One of the challenges of tera-scale visualization is that it is simply not possible to move "all the data" to the desktop for the purposes of visualization. Even if the local workstation were capable of storing and processing tera-scale data sets, movement of large data sets over the Internet is a tedious process. With the advent of high-speed testbed networks that are capable of gigabit speeds and beyond, new architectures for remote and distributed visualization are emerging, and Visapult is one such framework.

During the course of our work, we have performed field testing of Visapult on several platforms and using several high-speed testbed networks. During these tests, we instrumented the application using NetLogger, a subroutine callable library used to generate precision event logs in distributed applications. Based upon the data we obtained over the course of the project, we have effected improvements in the application, such as overlapping rendering and network i/o to increase overall efficiency. We were able to completely saturate every network link in all configurations we tested, including high-speed testbed networks such as NTON.

http://vis.lbl.gov/projects/visapult/

Internet2 Land Speed Record

TERRY GIBBONS, INFORMATION
SCIENCES INSTITUTE

A team from Microsoft, Qwest, the University of Washington and the Information Sciences Institute will demonstrate the transfer of data that won them the first Internet2 Land Speed Record award. The winning entry set a new standard for transcontinental Internet performance by transferring 8.4 gigabytes of data from Redmond, Washington, to Arlington, Virginia, (5,626 Km) in under 82 seconds. The rate of over 900 megabits per second is more than 15,000 times faster than a typical computer modem. Entries were judged on a combination of how much bandwidth they used and how much distance they covered end-to-end, using the standard Internet (TCP/IP) protocols.

Internet2 is a partnership led by over 175 US universities, working with industry, government and international partners. A primary goal of Internet2 is to deploy and demonstrate advanced networking capabilities that will make their way into the global commodity Internet. Internet2's Land Speed competition encourages research that focuses on how well the different network components work together rather than concentrating on increasing bandwidth.

The Land Speed Record highlights the need for dependable high end-to-end network throughput which is required by advanced applications but not commonly available to researchers today.

Details of the winning entries, as well as rules, submission guidelines and additional details are available at:

http://www.internet2.edu/html/i2lsr.html

High Definition TV

DAVID RICHARDSON, UNIVERSITY OF WASHINGTON

The University of Washington's "Internet HDTV" project provides a foundation for answering the question, "How would the world change if you could send studio-quality HDTV over a general-purpose Internet?" More fundamentally, it is intended to explore the intersection of network, video, and server technologies where near real-time distribution of extremely high-quality images is required.

In support of the ResearchChannel, UW's office of Computing and Communications has developed software capable of streaming studio-quality HDTV at speeds in excess of 200 Mbps. The project has been demonstrated over the Internet2 Abilene network, as well as across DARPA Supernets and temporary networks. Most recently, the system was used at the National Association of Broadcasters convention in Las Vegas to produce a news segment that originated in Seattle, was switched on the show floor, and returned for live broadcast to Seattle-area HDTV viewers.

A core principal of the Internet HDTV project is reliance upon standards from the video and networking communities. Broadcast industry codecs, data formats, and interfaces are used for origination and display. Commodity workstations, Gigabit Ethernet and Packet over SONET networking, and Internet Protocols are used for streaming. This has led to wider acceptance of the ideas than might have been expected, especially in a broadcast industry that equates the Internet with low-resolution video.

http://www.washington.edu/hdtv.

Virtual MechanoSynthesis (VMS)

JON GUICE, NASA AMES RESEARCH CENTER

The Virtual MechanoSynthesis (VMS) application allows users to see, move, and even feel simulated molecular structures in three dimensions. A user can grab an individual atom with a wand, move it about, and build it into arbitrary complex structures like Tinkertoys. The NAS data analysis group coupled an accurate molecular dynamics simulation code to an immersive graphical display with interactive capabilities and manual force feedback, allowing users to interact virtually with small collections of atoms "first hand." VMS uses computational steering—the ability to design and modify simulations interactively as they are running rather than relegating visualization to a post-processing phase—that allows users to immediately see the results of interactively changed parameters and provides opportunities to detect and explore patterns of cause and effect.

By running this simulation environment simultaneously in multiple locations, groups of scientists will be able to discuss their findings in real time. This demonstration is one in a series of demonstrations NREN is producing in collaboration with the NASA Astrobiology Institute.

http://www.nas.nasa.gov/Pubs/NASnews/1999/03/index.html

Keynote & Panel

5

Keynote Address, Raj Jain, Ohio State University/Nayna Networks

Panel: "Setting the Stage"

– John Wroclawski, Massachusetts Institute of Technology

– Kay Howell, National Coordination Office for Computing, Information and Communications

– Richard Carlson, Department of Energy

– Phillip Dykstra, WearOnEarth Communications, Inc.

– Ian Foster, Argonne National Laboratory

– Wesley Kaplow, Qwest Government Systems

Keynote Address
IP over Petabit DWDM Networks: Issues and Challenges

RAJ JAIN, OHIO STATE UNIVERSITY, NAYNA NETWORKS

In the last five years, dense wavelength division multiplexing networks have moved from research labs to deployment. Carriers are planning to offer multigigabit wavelength-on-demand services, and equipment vendors are designing optical switches that will handle over petabits of IP traffic in the core at a reasonable cost.

In this talk, we begin with recent DWDM records and present a sample of recent products and applications. Key technological developments that made DWDM possible will be explained.

Emergence of these optical DWDM networks has resulted in a very high-speed core within the Internet. Currently, the protocol stack consists of IP over PPP over SONET over DWDM. Industry is questioning the need for SONET in IP/DWDM networks.

Economic pressures are calling for the elimination of the SONET layer. What changes are required in IP protocols and in DWDM to make this possible? This is the topic of this talk.

Issues related to quality of service in such high-speed networks and recent developments in the area of multi-protocol lambda (wavelength) switching will also be discussed.

For further information, see:

http://www.cis.ohio-state.edu/~jain/

Setting the Stage Panelist

JOHN WROCLAWSKI, MASSACHUSETTS
INSTITUTE OF TECHNOLOGY

Setting the Stage Panelist - PITAC Perspectives on the Next Generation Internet

KAY HOWELL, NATIONAL COORDINATION OFFICE FOR COMPUTING, INFORMATION AND COMMUNICATIONS

Setting the Stage Panelist - Perspectives on Gigabit networks

RICHARD CARLSON, U.S. DEPARTMENT OF ENERGY

WHY WORK ON GIGABIT NETWORKING?

Communication is the corner stone of our modern civilization. From backyard gossip to cellular telephones, people communicate with each other. This peer-to-peer exchange of information unites individuals and helps to form communities. The ubiquitous PC, coupled to the Internet, has enhanced our ability to generate and share multimedia content (i.e., voice, video, data). As the amount of information content increases, the communications infrastructure needed to transport it must also increase. Fortunately, rapid technological advances, typified by Moore's Law, make this increased capacity both available and affordable. The question really becomes, "how can we not work on Gigabit networks?"

IS GIGABIT NETWORKING DIFFICULT, AND IF SO, WHY?

The basic underlying technologies for Gigabit networks are low-power CMOS electronic components, high-speed clocks and data transfer channels, and low-cost manufacturing facilities. In addition to the electronic components listed above, networks require low-cost optical components like solid-state lasers, fiber optic cables, and optical detectors. The ubiquitous low-cost PC running at gigahertz rates is just one example of what can be achieved by the electronics industry today. Modern fiber optic-based cross-country telephone networks running at OC-192 (10 Gbps) rates are an example of what the communications industry can achieve.

The challenge facing the data communications user comes in carrying today's demanding application flows. The Internet protocols and algorithms were not designed to operate over high bandwidth-delay product networks. These protocols need to be redesigned to provide better support for these types of networks without losing their essential congestion control features. The network interface card is still designed as an I/O peripheral device and thus does not work well when a continuous stream of packets arrives. This device needs to be redesigned to handle application level communication tasks without requiring constant service by the systems CPU.

In essence, we can easily build high-speed networks, but we have major problems building high-performance networks—that is, networks where applications can easily get a large fraction of the aggregate bandwidth that is available on the end-to-end path. Major research must be performed to identify and solve these problems.

PERSPECTIVE OF THE FUTURE OF GIGABIT NETWORKING AND TIMELINE FOR DEVELOPMENTS

Scientific applications can be viewed as precursors for what the general public will demand in the near future. The ubiquitous nature of the web, developed to support high-energy physics, is just one example of this. The current generation of scientific applications is being developed to support large, widely distributed communities of users. This will require the movement of terabytes of data from storage repositories to wherever the user happens to be. What the scientific community needs tomorrow, the entertainment industry will want in the next few years.

Setting the Stage Panelist - Issues Impacting Gigabit Networks: Why Don't Most Users Experience High Data Rates?

PHILLIP DYKSTRA, WAREONEARTH COMMUNICATIONS, INC.

The United States has several high-speed nationwide networks that support Research, Engineering, and Education. These networks should support data rates in excess of 100 Mbps, with many OC3 (155 Mbps), OC12 (622 Mbps), and even OC48 (2.4 Gbps) links. Routine end-to-end data rates approaching 100 Mbps is even a goal of the Next Generation Internet program. Yet most users today see perhaps one tenth of that goal. Why is this, and what should we do to improve the situation?

Recent measurements on the Defense Research and Engineering Network (DREN), vBNS, and Abilene networks have painted a rather grim picture of typical end-to-end performance. There appear to be many obstacles to high data rate flows. We briefly discuss some of them below, along with network design concepts that have become increasingly important as data rates have increased. Several of these concepts come directly from the estimate of TCP throughput:

*bps < min(rwin/rtt, MSS/(rtt*sqrt(loss)))*

WINDOW SIZE MATTERS

Most of our end systems still default to offering ~16KB TCP receive windows (rwin), or even 8KB. These values are fine for high-speed local area networks, and low-speed wide area networks, but they severely limit throughput on high-speed wide area networks. For example, over a coast-to-coast path (rtt = 40 msec), TCP could not exceed about 3.3 Mbps, even if it was running over a gigabit per second path.

The answer is not as simple as setting a large default rwin. Too large of a default window can run your system out of memory, since every connection will use it. A large window can also be bad for local area performance, and bad for some interactive sessions or applications that don't require high data rates. What is needed are tuned applications—ones that use large windows when and where appropriate—and/or adaptive TCP stacks that adjust rwin based on actual use. Web100 is an example of one project that aims to provide an adaptive TCP for the masses. We could do more today to improve typical user performance by improving end system software than we could by installing more high speed links.

LATENCY MATTERS

"High speed" networks are really high capacity networks. The "speed" with which data moves down a T1 line or an OC48 line is dictated by the speed of light in the media. Architectural changes in our high performance networks in the past few years have often resulted in increased delay or latency between pairs of sites. Examples include the relatively small number of Network Access Points (NAPs) where networks interconnect, the concentration of sites behind Gigapops, and the reliance on a fairly small number of very high capacity trunks.

On low capacity networks, latency caused by propagation delays wasn't a very critical factor. Today, however, we see numerous cases where the performance an application sees is directly impacted by the geographic path length of the network. Everything else being equal, TCP can go twice as fast if the path length (latency) is cut in half. Yet our routers today usually choose paths based on minimizing the number of hops, and following the highest capacity path, even if that means routing across the country and back. Many applications will do better over a low latency OC3 path than over a high latency OC48 path, yet we have no way to ask for such a path from the network.

Setting the Stage Panelist - Issues Impacting Gigabit Networks: Why Don't Most Users Experience High Data Rates? cont.

PHILLIP DYKSTRA, WAREONEARTH COMMUNICATIONS, INC.

On a single high performance network today, measured latencies are typically ~1.5x - 3x that expected from the speed of light in fiber. This is mostly due to taking longer than line-of-site paths. Between different networks (via NAPs) latency is usually much worse. Some extra distance is required, based on the availability of fiber routes and interconnects, but much more attention should be given to minimizing latency as we design our network topologies and routing.

MTU MATTERS

Packet size can have a major impact on throughput. The dynamics of TCP are such that, for a given latency and loss rate, there is a maximum packet per second rate that can be achieved. To increase throughput, you have to increase the packet size (or reduce latency or loss, which is something the end systems can't control).

Today the world is rapidly heading to where 1500 bytes is the largest supported end-to-end packet size. This is because of the dominance of Ethernet technology and the use of 1500 bytes even at gigabit data rates. Such small packets are a major obstacle to high performance TCP flows. At one gigabit per second, this equates to over 83000 packets per second, or only 12 microseconds per packet. There is no reason to require such small packets at gigabit data rates.

In the short term, the author hopes that the 8KB "jumbo frame" proposal for gigabit Ethernet becomes widespread. In the longer term, we should build high speed networks that can support much larger packet sizes. The backbone links and NAPs are particularly important, because if they restrict MTU, the end systems are helpless. It is hard to overstate the importance of this issue.

LOSS MATTERS

In the old days we thought 10% packet loss was acceptable. After all, TCP does error recovery, and 90% isn't bad, right? Today, many service level agreements (SLAs) target a loss of 1% or less (often averaged over 24 hours). For gigabit data rates, however, loss has to be astronomically small!

For example, to achieve a gigabit per second with TCP on a coast-to-coast path (rtt = 40 msec), with 1500-byte packets, the loss rate cannot exceed 8.5x10^-8! If the loss rate was even 0.1% (far better than most SLAs), TCP would be limited to just over 9 Mbps. [Note that large packet sizes help. If packets were n times larger, the same throughput could be achieved with n^2 times as much packet loss.]

Gigabit networks thus need to be nearly lossless. We believe that one of the reasons that such low loss isn't being observed is because most of today's routers and switches have insufficient buffering for such high bandwidth-delay products. Few of the high performance paths we have studied show stable queuing regions. Also, the concept of depending on loss to indicate congestion to TCP may not apply very well at extreme bandwidth-delay products.

BUGS ARE EVERYWHERE (ESPECIALLY DUPLEX ONES)

Recent measurements over numerous high performance paths have turned up a wealth of bad behavior, much of which is still unexplained. Slow forwarders, insufficient buffering, strange rate shaping behavior, duplex problems, packet reordering, and low level link or hardware errors are all playing a part. Very few paths can sustain the packet per second data rates that you would expect from the underlying hardware and links.

Setting the Stage Panelist - Issues Impacting Gigabit Networks: Why Don't Most Users Experience High Data Rates? cont.

PHILLIP DYKSTRA, WAREONEARTH COMMUNICATIONS, INC.

At least one problem deserves special mention. The failure of Ethernet auto-negotiation, and the resulting duplex problems, are perhaps the single biggest performance bug on the Internet today. The results of this bug only show up under load that makes them difficult to notice. Low rate pings show almost no loss, but high data rate loads result in dramatic loss. Our tests, and similar reports from others, are indicating that this bug has reached epidemic proportions.

WHY DEBUGGING IS HARD

Network test platforms and programs are usually only available at the edges of the network. When end-to-end tests are performed, they span many different devices and links. The result is a messy convolution of all of the behaviors along the path. When bad behavior is observed, it is sometimes nearly impossible to figure out where in the path the problem lies.

Performance problem debugging would be vastly easier if routers provided some kind of high performance testing service. Routers are designed to forward packets well, but are usually very bad at answering traffic directed to them. This means that tests can't be directed at a router in order to debug a path problem hop-by-hop. If you target a router in the middle of the path, you get such poor performance that other problems you are looking for are usually masked. The participation of routers in something like the proposed IP Measurement Protocol (IPMP), and/or a high speed echo service, would greatly aid in debugging.

SECURITY ISN'T HELPING

The ever increasing security threat, and level of abuse on the Internet, has led to numerous measures that decrease performance and make performance measurement and debugging more difficult. ICMP is often blocked making ping and/or traceroute impossible. Deliberate rate limits are sometimes imposed on ICMP or other traffic as a measure to defend against denial of service attacks. Sometimes only a limited number of TCP and UDP port numbers are left unblocked, which can prohibit measurement applications that use other ports. And an increasing number of Firewall and Network Address Translation (NAT) boxes are in the path, creating a loss of end-to-end transparency.

The performance impact of all of these measures has not been well studied. What exactly is the slowdown of different routers given certain kinds of filter lists? How fast do various firewall and NAT devices forward packets under different traffic situations? Can you bypass these security mechanisms for authenticated applications? The use of ICMP for measurements should probably be phased out, but acceptable alternatives to ICMP need to be created.

MEASUREMENTS ARE EASY, ANALYSIS IS HARD

We are doing well today at collecting basic measurements. Projects like AMP, Surveyor, PingER, and RIPE, are gathering a wealth of delay, loss, and route information. What we aren't very good at yet is learning things from all of that data. Major progress could be made from detailed automated analysis of the data: detection of anomalies, correlation of events, high level abstraction of causes. There are several projects working in this direction, but we are just beginning.

Setting the Stage Panelist - From Networks to Grids: The Need for Innovation in Infrastructure

IAN FOSTER, ARGONNE NATIONAL LABORATORY AND THE UNIVERSITY OF CHICAGO

Over the past several decades, advances in science and engineering have motivated—and been enabled by—continued innovation in information technology infrastructure. Examples of such innovations include supercomputers and supercomputer centers as well as the Internet. Today's research and education networks are the latest instantiation of this trend.

We currently face, once again, major changes in the nature of science. Specifically, advances in simulation science and instrumentation and an increased emphasis on multidisciplinary, distributed collaboration (powered in part by the emergence of the Web) combine to create a situation in which large communities can cooperate in the creation and analysis of large bodies of data. Numerous prototypes and experiments have demonstrated the feasibility and utility of new problem-solving approaches based, for example, on secure remote access to online instruments, distance collaboration, shared petabyte datasets, and large-scale distributed computation.

These new modalities of scientific enquiry require, in many cases, high-speed networks. But the central problem faced by developers is not high-speed transport but rather *resource sharing and coordinated resource use in dynamic, multi-institutional "virtual organizations."* This central problem is not adequately addressed by existing infrastructures, falling as it does outside the traditional purviews of both individual sites and network service providers. Yet a lack of suitable infrastructure is significantly impeding our ability to realize new "net science" modalities.

I believe that the solution to this problem is the creation of a new class of infrastructure focused on enabling resource sharing. This infrastructure must support new methods for establishing identity, defining policies, and negotiating access in dynamic, multi-institutional settings. It must include resource discovery protocols and services that allow small and large groups to publicize resource availabilities, as well as resource management protocols and services that allow those groups to coordinate how resources are allocated. Other requirements include new protocols and services for accessing remote end system resources, for co-allocating and integrating those resources to achieve end-to-end performance goals, for high-speed data movement, and for monitoring the performance of distributed applications.

The term "The Grid" has gained some currency as a descriptive term for this new class of infrastructure. Various R&D projects, including the Globus project in which I am involved, are developing "Grid Services" that meet some of the requirements just listed. Major scientific and engineering consortia are deploying such services on a large scale, in effect building the first Grid infrastructures: for example, NASA's Information Power Grid, NSF's National Technology Grid, DOE's DISCOM, and the European high energy physics community. Our understanding of what it means to build "Grids" will evolve, but there is certainly an experience base on which to build.

In conclusion, I argue that any future technology roadmap for Gigabit Networking must aggressively embrace this larger vision of what it means to provide useful infrastructure, moving beyond conventional notions of network and end system to define and build integrated Grids.

[For more info, see "The Grid: Blueprint for a Future Computing Infrastructure" (I. Foster, C. Kesselman, Eds, Morgan-Kaufman, 1999), also Grid Forum (www.gridforum.org). For Globus, see www.globus.org. The HPDC conference (www.hpdc.org) brings together Grid researchers.]

Setting the Stage Panelist

WESLEY K. KAPLOW, QWEST GOVERNMENT SYSTEMS

It was not too long ago that 64 Kbps and 1.5 Mbps private lines dominated private networking, and public data networks were constructed using 45 Mbps circuits. Today, the picture has changed dramatically. Most private networks have transitioned to using public data networks, with access rates from 1 Mbps all the way up to 622 Mbps and more. These public networks are constructed using 2.5 Gbps and 10 Gbps links.

This tremendous capacity has been enabled by the rapid development of opto-electronics, and by the deployment of optical fiber throughout the United States as well as other locations such as Europe. The combination has led to the emergence of new carriers that have the ability to put 500 Gbps over a single fiber using Dense Wavelength Division Techniques (DWDM), enabling the use of high capacity links for various types of public data networks using ATM and Packet-over-SONET technology.

The current conventional architecture is to form SONET rings using high-capacity fiber transmission systems to meet the high-availability requirements demanded by customers. However, as the number of DWDM channels increases, other architectures that use OSI layer 2 and 3 re-routing techniques to ensure high-availability are being considered. This is leading to a rethinking of the architecture of public data networks, both as a cost reduction method (i.e., eliminating the SONET ring protection), and as a method to reduce the complexity and time required for provisioning new bandwidth. Cost reduction, and the ability to quickly meet rising customer bandwidth requirements are necessary to ensure that a carrier meets the demands of the emerging competitive marketplace for Megabit and Gigabit bandwidth.

The private use of public data networks to obtain Mbps and a low number of Gbps capacity is likely to be the dominant method for some time. However, there is another service that is now becoming available. These are the so-called "wavelength" services that provide multiples of 2.5 Gbps or 10 Gbps unprotected wavelengths from source to destination. This means that the bandwidth previously available only to a facilities-based network provider can now be provided directly to a customer. So, from a customer's point-of-view, there is an increasing number of options to obtain Gbps of network capacity. Facilities-based service providers are creating an infrastructure that can scale capacity between public routers and switches while at the same time providing this capacity directly to the end-user.

However, as much as this seems to be positive for Gigabit networking, here are a few words for caution. Even the most robust and established public data networks have only tens of Gbps of total network capacity, and are based on current Internet Protocols that have evolved during a time of Kbps and Mbps network capacities. Issues such as route instabilities and re-route time due to failure will have a much larger impact on customer applications as the amount of data "in transit" becomes tens of Megabytes instead of Kilobytes. Moreover, the expected use of public data networks in the Internet world is to provide for tens of thousands of individual short-lived data streams that aggregate to Gbps, not few long-lived flows each requiring Gbps. It is not clear what impact Gbps requirements will have on the planning and operations of these networks. For customers that choose public networking, customer applications and their supporting middleware will have to be able to adapt to Gbps re-routes, latency changes, potential duplication and out-of-sequence data. Customers that choose the "wavelength" approach will have similar problems, but will also have to contend with additional local access requirements, as well as core router hardware selection and operational procedures.

Technology Presentations

Developing the Infrastructure

– Jim Gimlett, Network Elements: Platforms

– Leonid Kazovsky, Stanford University: Optical Access

– Nick McKeown, Stanford University: Fast Routing and Switching

– Basil Irwin, University Corporation for Atmospheric Research, & Matt Mathis, National Laboratory for Applied Network Research: Web100

Tools for Gigabit Networking

MEASUREMENT

– Joe Evans, University of Kansas

GIGABIT END-TO-END ISSUES PANEL

– Steve Corbato, Internet2: Overview

– Michael O'Connor, Brookhaven National Laboratory: Campus

– Cas D'Angelo, Southern Crossroads Gigapop: Gigapops

– Jerry Sobieski, Mid Atlantic Crossroads: Peering

– William Lennon, Lawrence Livermore National Laboratory: Wide Area

Developing the Infrastructure - Next Generation Platforms

JIM GIMLETT, NETWORK ELEMENTS

Advances in DWDM transport, Terabit Switch Routers, optical switching, Gigabit Ethernet and other breakthrough technologies have resulted in skyrocketing WAN and LAN bandwidths and performance. Similarly, advances in microprocessors, memory bandwidth, storage and interconnect technologies have brought about stunning improvements in networked computing capabilities. These gains enable us to entertain the vision of multigigabit distributed applications, such as distributed computing, sensor fusion, data mining, and telepresence, requiring high bandwidth, secure and transparent access to sensor, processing and storage resources. Realizing this vision also requires host platforms to be able to usefully ingest and output multigigabit streams of data to other platforms and information nodes regardless of where they are located. Today, this is difficult to accomplish even for resources contained within the same box. Despite the impressive gains in networking and computing technologies, the challenge of seamless access to information resources at multigigabit rates by the end user or application remains an elusive goal for all but the most advanced supercomputing platforms.

So what has happened? Certainly, this problem has not been ignored. Achieving gigabits/sec to the end user was one of the goals of the Gigabit Testbed Initiative program of the early '90s. The CNRI's Gigabit Testbed Final Report of 1996 stated that: "Host I/O was one of the most challenging areas of the testbed effort. In general, it proved to be the Achilles' heel of gigabit networking—whereas LAN and wide area networking technologies could be and were operated in the gigabit regime, many obstacles impeded achieving gigabit flows into and out of the host computers used in the testbeds."

Today, in this new decade and century, we now have a handful of "hero" demonstrations of gigabit flows to the end application. These tend to be simple demonstrations of FTP file transfers, with most of the host processor's cycles spent in processing the FTP/TCP/IP protocol stack. (Don't even think about asking to perform real-time encryption/decryption on that flow.) End user access to multigigabit isochronous data flows is a much tougher problem to crack. So although we have made some progress, bottlenecks in the platform remain one of the toughest challenges of gigabit networking.

In this talk we will examine some of the remaining platform bottlenecks and discuss a few of the initiatives in platform architecture and design likely to bear fruit in the next couple of years. We will also describe a project with DARPA designed to facilitate distributed host platform I/O access to gigabits/sec sensor data. Finally, we will try to look further out in asking, what is a platform? what is the difference between a platform and a network element? and where does the platform end and the network begin?

Developing the Infrastructure - Perspectives on Future Optical Networks

LEONID KAZOVSKY, STANFORD UNIVERSITY

Photonics plays an increasingly paramount role in high-speed networks, from the backbone to access networks. With the development of new device technology, photonics continues to penetrate/integrate into network electronics, moving closer to the end user. Currently, photonics provide little complexity when compared to their electronic counterparts. Therefore, a merge must occur between the two that optimally combines the strengths of both technologies. In addition, much of the bandwidth available on optical fiber has yet to be utilized. New optical devices, improved optical amplifiers, and promising fiber technologies will allow great increases in capacity as new methods for capitalizing on this available bandwidth are developed.

As backbone traffic reaches higher bit rates, network complexity will be pushed toward the edge of the network. High capacity point-to-point links have long been the focus of optical communications research, and as such, are well developed. Metro Area and Access Networks will play an important role in tomorrow's networks, and research efforts have grown in this area as a result. The Optical Communications Research Lab (OCRL) at Stanford University has worked on Metropolitan Area Network (MAN) projects in both the circuit-switched and packet-switched arenas, exploring such topics as passive optical networks, optical contention, optical packet switching, and reconfigurability.

Micro-electromechanical systems (MEMS) technology and other recently developed optical cross-connect techniques will rapidly replace other protection-switching mechanisms. As optical switches become faster and more reliable, they will play an important role in provisioning and reconfigurability. Other new devices like ultra-fast tunable transmitters and tunable optical filters allow for network topologies based on wavelength routing. The OCRL has realized novel MANs through the use of such new devices. Attempts to push network control towards photonics will continue, limited only by the vast difference in complexity offered by photonics when compared to electronics. An all-optical IP router, for example, is quite unrealistic at this point in time, while networks which utilize optical devices to route data streams based on which wavelength they are carried exist today, and are quite promising.

Research on fibers, optical amplifiers, modulation techniques, and other areas will result in new ways to exploit the available bandwidth on a fiber. The combination of a new fiber that suppresses the water absorption peak with ultra-wideband optical amplifiers will allow for >50 THz of bandwidth on a single fiber. The OCRL at Stanford University actively researches fiber nonlinearities and ultra-wideband optical amplification. They have demonstrated techniques that amplify over 200nm of bandwidth using optical parametric amplification.

Most current transmission systems employ wavelength-division-multiplexing (WDM) technology to access the bandwidth afforded to them by their optical amplifiers. Each wavelength has a data rate dictated by either the transmitter/receiver drive electronics or by system nonlinearities and device limitations. As electronics become faster, WDM devices improve, and techniques are developed to combat fiber nonlinearities, channel spacing will decrease and channel rates will increase.

Research in photonics is driven, in large part, by the seemingly insatiable demand for bandwidth. Fiber optics is capable of meeting such demands—therefore, it will move towards the end user as demand for bandwidth increases. Optics dominates the backbone and MAN space; enterprise and local area networks employ optics as well. Photonics is the means by which gigabits-per-second will make it to the end user.

Developing the Infrastructure - Fast Routing and Switching

NICK MCKEOWN, STANFORD UNIVERSITY

Developing the Infrastructure - Web 100

BASIL IRWIN, UNIVERSITY CORPORATION FOR ATMOSPHERIC RESEARCH
MATTHEW MATHIS, NATIONAL LABORATORY FOR APPLIED NETWORKING RESEARCH

Tools for Gigabit Networking - Measurement in Gigabit Networks

JOSEPH B. EVANS, UNIVERSITY OF KANSAS, LAWRENCE

When new high speed networks are first deployed and made available to users, experience has shown that performance often falls below expectations. The reasons for this are varied, ranging from end-system performance to protocol sensitivities to packet loss in the network. Measurement infrastructure is necessary to determine the cause of these bottlenecks and eliminate them.

The types of measurements to be performed are driven by the dominant traffic types in the network under consideration. In very general terms, traffic in a gigabit network might be a single or small number of flows of high volume, or traffic might be the aggregate of many flows that make up a high speed whole. In the former case, test applications are often first applied to exercise the network, and the measurements are used to fix bottlenecks. In the latter case, measurements are more often used to adjust traffic patterns to route around bottleneck areas. In either case, measurement information is necessary to make informed decisions.

For single flows and constrained networks, manual configuration of measurement points may be feasible. As the number of flows and complexity of the network grows, this is rarely reasonable due to the varied patterns and many network elements involved. This observation helps motivate research on flexible measurement infrastructure for high speed networks.

It is desirable to enable applications to easily make network performance decisions, should such an infrastructure exist. In addition to an adaptive monitoring infrastructure, the capability to archive, publish, and analyze the data is necessary. A network options recommendation service might be a useful mechanism to provide applications with the necessary information.

A rich set of features is desirable in the ideal network monitoring and adaptation system. Support for measurements on both end systems and networks is necessary, as is support for measurements spanning multiple networks. An important requirement is support for correlation of events such as packet loss and application throughput and delays. Support for measurements with various time granularities is needed due to the limitations of the probes. Robustness to network failures and measurement probe errors is required, since these are the very events that may be leading to less than expected performance. It is desirable that the system be scalable to many distributed applications and be simple to use by application users. Aids for identifying, debugging, and responding to anomalous conditions are also desirable. Several pieces to building this puzzle will be described next.

There are a number of tools that are being used in current high performance networks. A very useful toolkit developed at Lawrence Berkeley National Laboratory (LBNL) is based on NetLogger, a standard event logging format. This toolkit provides methods for monitoring hosts and applications and tools to visualize logs. In order to initiate measurements at the proper times, the Java Agents for Monitoring and Management (JAMM) tool was also developed at LBNL. This monitors specified ports on any host, and when traffic is detected, starts up NetLogger tools. The NetArchive toolkit was developed at the University of Kansas (KU) to support data collection using application hooks and SNMP. This provides for data archiving using a meta database and time series database in order to support the large volumes of data collected at reasonable time resolutions in high speed networks. The records in this system are stored in NetLogger format. Several other tools, such as NetSpec, a tool developed at KU for controlled traffic generation, and a modified version of tcpdump, have been enhanced with Netlogger capability. A future addition to this collection will likely be a network advice service that analyzes the vast amount of monitoring data to provide advice to applications on the best use of the network.

The bits of measurement infrastructure described above are being developed in ongoing research and deployed in testbeds in order to enhance the performance of high speed applications.

Tools for Gigabit Networking - Gigabit End-to-End Issues Overview

STEVE CORBATO, INTERNET2

Tools for Gigabit Networking - Gigabit End-to-End Issues
Campus Issues

MICHAEL O'CONNOR, BROOKHAVEN NATIONAL LABORATORY, NEW YORK

CURRENT STATUS

Gigabit Ethernet made its debut at Brookhaven National Laboratory (BNL) on the Relativistic Heavy Ion Collider project (RHIC) in 1997. Since that time Gigabit Ethernet has gone from niche to mainstream campus deployment. BNL has adopted Gigabit Layer 3 switches as its campus core routing engine, migrating from an ATM LANE "router on a stick" design. While not yet deployed at the desktop, Gigabit Ethernet is the uplink of choice for new layer two switch upgrade projects.

Jumbo frames were a distinct advantage two years ago and enabled RHIC data acquisition servers to achieve target data rates to the central computing facility, but as processor speed increases, the strong case for jumbo frames at RHIC erodes.

Lacking an external single-mode fiber pair, many campus buildings are forced to invest in new fiber before they can take advantage of Gigabit Ethernet connectivity back to the collapsed backbone. However, Gigabit Ethernet is used inside a building where multi-mode fiber is plentiful and well suited to span the short distance between wiring closets.

Marrying ATM ELANs to 802.1Q/ISL VLANS has been essential to an orderly migration from an ATM core to a collapsed layer 3 Ethernet switch backbone. While fairly convoluted, it is working and has allowed BNL to extend the useful life of our ATM layer two-core mesh.

WHAT THE FUTURE HOLDS

Gigabit Ethernet has accelerated the deprecation of several technologies at BNL. ATM networking equipment, software-based routers and non-IP network protocols are all gone from future network plans. Gigabit layer three switches optimized for IP are replacing software-based routers that supported many more protocols than do the current generation of high performance firmware-based routing switches. Not only are the legacy routers disappearing, so are non-IP legacy protocols and 10Mbps Ethernet links in the core/backbone fading toward distant memory. ATM did fulfil mid-nineties bandwidth requirements, but at a cost and complexity way beyond that of newer Gigabit Ethernet technology that has greater market support, vendor diversity and a clearer future in the campus LAN.

BNL is not alone with network security looming large in the collective consciousness of the enterprise. Gigabit data rates present a whole host of performance-related issues with respect to various elements of a cyber security system. Access control firewalls and packet sniffing intrusion detection systems lag the gigabit technology curve and will have to catch up to adequately handle their respective roles in the security architecture.

There is now enough bandwidth in the campus to begin serious discussion of delay sensitive applications such as voice and video.

http://www.itd.bnl.gov/nss

Tools for Gigabit Networking - Gigabit End-to-End Issues
Gigabit Networking Issues Related to GigaPOPs

CAS D'ANGELO, GEORGIA INSTITUTE OF TECHNOLOGY

Gigabit Ethernet is a well-accepted technology throughout the networking industry. Compared to technologies like ATM and POS, it is inexpensive in terms of acquisition and operation. Attracting qualified staff to operate gigabit Ethernet equipment is easier than ATM and POS. Gigabit Ethernet will be a player in the gigaPOP arena, but will not fit all models.

Currently at SoX (Southern Crossroads GigaPOP), we operate across an ATM backbone. The core of this is a distributed set of three ATM switches. Through two locations, SoX attaches customers and networks and uses its ATM infrastructure to attach to a central router.

SoX is likely to begin using VBR (variable bit rate) PVCs (permanent virtual circuits) to provide services across the same circuit, with different qualities of service. For instance, a university could purchase a DS-3 circuit into SoX (it could be SONET service or ATM service). The circuit could then be divided into two different pipes. One circuit would be for Internet2 service. It would be classified as VBR and would have a maximum available bandwidth that is guaranteed (perhaps 30 Mbps).

The other circuit would be for ISP services (Internet1). It would be classified as UBR (unspecified bit rate). It would not be able to preempt the VBR Internet2 traffic, but would be able to use any unused bandwidth that was available. This allows it to use the entire 45-Mbps pipe in the event of zero Internet2 traffic.

If the Internet2 pipe were operating at its full capacity (30 Mbps), the Internet1 traffic would be restricted to the remaining bandwidth (15 Mbps).

This type of service differentiation is early in development for gigabit Ethernet. It will take strong, interoperable protocols like MPLS (multi-path label swapping) to provide this type of differentiation.

Another feature of ATM is its ability to be purchased from telcos. For SoX, customers are able to buy either a SONET or ATM service to one of our gigaPOP locations.

Gigabit Ethernet service is more difficult to find and will likely be provisioned as a wavelength across a DWDM (dense wave division multiplexing) network. Bits per dollar, purchasing wavelengths is likely to become more attractive as prices begin to drop for long-haul circuits.

SoX and SURA (Southeastern Universities Research Association) are looking into possibilities of procuring rights to fiber across the region. With a DWDM network laid over this fiber, it will allow SoX to provide a gigabit Ethernet per service if necessary and leave leftover wavelengths to use for experimental networks.

In the event SURA is successful, it would be likely that SoX would implement a gigabit Ethernet backbone and use overprovisioning as our QoS for the short term. The gigabit Ethernet equipment is cheaper than its equivalent OC-12 interfaces.

Universities are turning to gigabit Ethernet as an alternative to ATM for high-speed network access. Because many of the QoS and convergence technologies have not been implemented on university campuses, cheaper and faster gigabit Ethernet is a great alternative. The standardization of 802.1q for VLANs and 802.1p for class of service lets gigabit Ethernet approximate the features that the more expensive ATM is being used for. Services like multicast are designed for broadcast media and have been easier for vendor designers to implement on gigabit Ethernet.

Tools for Gigabit Networking - Gigabit End-to-End Issues
Gigabit Networking Issues Related to GigaPOPs cont.

CAS D'ANGELO, GEORGIA INSTITUTE OF TECHNOLOGY

In the gigaPOPs, ATM is still entrenched because of the ease and pricing of the SONET and ATM services to be ordered form the telcos. Because ATM can be ordered in many different speeds, it allows for purchasing just the increment of bandwidth one needs. In addition, ATM allows easy flexibility in constructing different pipes for different services.

As a networking manager, I have had to hire several positions in the past few years. Very few candidates have been able to spell ATM. All of the network specialists working on our ATM infrastructure have learned through our team. On the other hand, people experienced with Ethernet and gigabit Ethernet have been easier to find and attract (although still not trivial). This provides us with a reason to try to develop ATM-to-gigabit Ethernet migration programs.

It will take some time before gigabit Ethernet becomes fully adopted in the gigaPOP. As soon as gigabit Ethernet services become available from telcos or wavelengths become less expensive, gigabit Ethernet will begin to take off. As MPLS develops, the QoS concerns will begin to be addressed. From a gigaPOP perspective, gigabit Ethernet is attractive because of ease of management and inexpensive equipment.

http://truckasaurus.ns.gatech.edu/casman/

Tools For Gigabit Networking - Gigabit End-to-End Issues Peering

JERRY SOBIESKI, MID ATLANTIC CROSSROADS

Tools For Gigabit Networking - Gigabit End-to-End Issues Wide Area

WILLIAM LENNON, LAWRENCE LIVERMORE NATIONAL LABORATORY

Technology Roadmaps

7

Gigabit Testbeds

Platforms

Measurement

Middleware/Integration

Gigabit Testbed Roadmap

1 INTRODUCTION

The objective of the Gigabit Testbed breakout session was to establish a roadmap describing how to proceed with gigabit testbeds until the end of the NGI program in September 2002 and discussing how testbeds should evolve post-NGI. The session focused on knowledge and technology transfer rather than on research issues.

2 STATUS— CURRENT GIGABIT NETWORKING ACCESS

2.1 Internet2/Abilene Network

Abilene is a research testbed for advanced services and applications providing a national education intranet interconnecting all U.S. educational institutions to enable applications and services unavailable over the commercial Internet. Current Internet2 architecture consists of multiple backbones to which campuses can connect either directly or through gigapops.

Transit connections support both IP over SONET (OC12 and 48) and ATM (OC3 and 12). Abilene serves as an interconnection for peering with federal and international R&E networks. Abilene currently has 40 sites directly connected with 10 in progress. A majority of universities participate, collaborating on shared connectivity to Abilene via gigapops.

For more information, see *www.ucaid.edu/abilene*.

2.2 SuperNet: DARPA NGI Research Network

The DARPA SuperNet includes ATDnet/MONET, NTON II, BOSSnet and ONRAMP with HSCC as the connecting OC48 cloud.

Twenty-nine institutions are directly connected to the SuperNet while others are connected via tunnels. Forty-five percent of the institutions currently connected can support gigabit rates to the desktop. That number is expected to reach 62 percent by the end of 2000. Seventy-three percent of the institutions currently connected can support 500 Mbps to the desktop, and 41 percent currently have dark fiber.

For more information, see *www.ngi-supernet.org*.

3 ISSUES AND CHALLENGES

In order to support NGI goals economically, we need to educate people as to what is possible and how to get to gigabit levels. We also need to close the gap between the network people and the applications people. The key to closing the gap is technology transfer and increasing user/community awareness in the campus and operations arenas.

3.1 Gigabit Networking Access

The next step for many campuses would be to upgrade their WAN access link from OC3 to OC12, yet the cost of OC12 is prohibitive for most of these campuses. What is needed is access to dark fiber. The growth area for Abilene and the gigapops now involves many smaller clients upgrading their access links from T1 to DS3.

3.2 Network Issues

3.2.1 Local and Campus Area

More institutions in the local and campus area must become involved. The question is how to make it easy for campuses to upgrade their connections. One suggestion was to form a “gigabit strike force” to aid in the process.

In the areas of jumbo frame support, gigabit to the desktop architectures, commodity/research traffic separation strategies and scaling of high bandwidth application flows into more commodity campus traffic, the following recommendation was made:

Gigabit Testbed Roadmap cont.

3.2.1 Local and Campus Area cont.

Recommendation: DARPA, Internet2 and NSF should identify a team whose charter is to select campuses and/or create architectures that demonstrate best practices in the above areas.

3.2.2 Wide Area

On the wide area scene, the following issues were identified:

- Economic viability of high bandwidth institutional connections: How do we support these connections?
- Testbed persistence: OC48 links are gone when applications are ready to use them.
- Pushing gigabit links into broader community
- Evolution of the NGIXs: OC12 ATM interconnects at the NGIXs are a hard limit.
- Peering: Splitting out Internet2 commodity and high bandwidth traffic, v4 tunnels
- Jumbo frame support
- ATM/POS integration

Recommendation: These issues are best addressed within the JET working group. Many of the issues such as persistence and expansion are directly related to funding issues best addressed by the funding agencies.

3.3 Workstation Issues

Regarding workstation issues, the following elements were identified:

- Report from Platforms Breakout Session
- Jumbo frame support
- Disk I/O
- TCP buffer management

Recommendation: Defer to the recommendations of the Platforms Breakout Session.

4 RECOMMENDATIONS

4.1 Enhance TCP

Enhance TCP to enable 500 Mbps+ flows (Web100 Project).

4.2 Couple Application Researchers and Network Researchers

Provide closer coupling of applications researchers, and network researchers and testbeds at both the technical and higher layers.

Recommendation: HPNAT and JET hold “Magic 10” workshop in late 2000 or early 2001 to ensure applications have the network support needed.

4.3 Continue Emphasis on Technology Transfer

Continue emphasis on technology transfer to commercial sector to help ensure that NGI institutions drive (at some level) new service models in the commercial sector.

Example: Commit resources necessary to begin design and implement All Optical Exchanges (AOXs) to replace NGIXs in the next five years.

4.4 Challenge the Internet2/Abilene Community

Throw out challenges to the Internet2/Abilene community to encourage people to develop applications.

Recommendations/suggestions:

- Publish selection of Magic 10
- Internet2 Land Speed Record
- SCInet SC-2000 Network Challenge
- Internet2 Application Excellence

Gigabit Testbed Roadmap cont.

4.5 Make it Easy

- Provide a knowledge base: Document and publish lessons learned in solving gigabit end-to-end problems.
- Increase support to application developers to allow the network to become more transparent.
- Form a “Strike Force” to provide one-stop shopping for applications developers (as opposed to campus, gigapop, WAN networks) to ensure applications developers have a well-known network POC.

Recommendation: Establish a Gigabit Network Strike Force (a focused team available for assisting applications researchers with wide area/platform gigabit issues).

4.6 Provide for Testbed Persistence

Obtain commitments to provide for ongoing investment in testbeds in the future.

4.7 List 10 Applications

Make a list of candidate applications to meet the goal of 10 gigabit applications by 2002.

5 EVOLUTION CONSIDERATIONS

Questions to consider for the near term:

- How many institutions are connected at gigabit speeds, OC-12 or higher?
- Are connections supported on a persistent basis?
- What technologies are emerging?
- What are the “Magic 10” applications, e.g., broadband video applications (HDTV, Digital Amphitheater) point-to-point and multicast?

Recommendation: HPNAT to identify the Magic 10 by the end of this year.

Platforms Roadmap

1 INTRODUCTION

The first steps towards drafting a technology roadmap include defining the scope of the subject, determining the current state of the technology available, and ascertaining the desired destination.

A non-trivial amount of time was spent agreeing on what was meant by the term “platform” as it pertains to gigabit networking. The advent of distributed technologies such as grids has clouded the picture. In the end, “platform” was defined as ending where the network begins—at the network interface card—which is considered to be within both platform and network. Platform was further defined as having two major constituent parts, hardware and software, each meriting its own discussion.

2 STATUS

2.1 Current Status

The workshop application demonstrations showed that end-to-end gigabit networking is possible today, but that it requires significant advanced planning, demonstration-specific networking configurations, and extensive testing, tuning and tweaking of both end system and network equipment by staff with “guru/wizard” level expertise.

2.2 Near-Term Status

The roadmap goal for gigabit networking platforms at the end of the NGI program is to be at the point where someone with an intermediate level of technical ability at a facility with access to appropriate connectivity can implement a gigabit application using off-the-shelf equipment without significant configuration development and performance stuning.

2.3 Hardware

Of the two major constituent parts of gigabit networking platforms, hardware is in fairly good shape. Gigabit Ethernet interfaces are already commodity items, and it is expected that pre-standard 10-gigabit Ethernet interfaces will begin to appear within two years. Today one can buy a commodity computer with a gigaflop CPU and 10/100/1000 network interface standard for less than $2000. There is still room for improvement, however, as such things as bus bandwidth, design implementation, and firmware revisions can have a drastic impact on hardware performance.

2.4 Software

By comparison with hardware, there is definite room for improvement within the software arena as it pertains to gigabit networking. This is true both within the network stack and the operating system. Not every OS will allow users to tune features, and not every TCP stack has capabilities such as SACK, ECN, etc., useful for a gigabit networking platform.

3 ISSUES AND CHALLENGES

Significant tuning and tweaking of obscure hardware, OS, and TCP parameters are currently required to enable gigabit applications. This is one of the biggest obstacles as a thorough understanding of relatively arcane issues is often required to get things to work. It is also very difficult to determine if required capabilities are available within a particular product without hands-on evaluation.

Platforms Roadmap cont.

4 RECOMMENDATIONS

NGI and Internet2 entities can ease the path to ubiquitous gigabit networking through a combination of research, documentation, education, and leadership through example, specifically:

– Performing and supporting research necessary to understand the role and relevance of networking parameters and advanced networking capabilities such as auto-negotiation and adaptive tuning of TCP parameters.

– Implementing and widely distributing a "gigabit networking friendly" operating system distribution containing capabilities like an instrumented TCP stack, SACK, adaptive window sizes, auto-tuning host MTU size to take advantage of jumbo frames where available, etc., to inspire commercial OS developers to include these features in future releases.

– Providing a way for prospective gigabit networkers to determine which products would be suitable platform elements without hands-on evaluation. Although a list of tested/approved equipment may be a good start, the market moves so fast that any such list may become obsolete before it can be widely distributed. Another way to meet the same goal would be to create a checklist of features relevant to gigabit networking so that manufacturers know what they need to include in their product in order to obtain the "Good Networking Performance Stamp of Approval."

– Creating a set of benchmarking utilities relevant to gigabit networking platform elements and/or hold a gigabit networking performance "bake-off" competition so that there is a way to differentiate among various products.

– Documenting and distributing knowledge already assembled so applications people do not need to become network experts. A good model may be that of the TCP and satellite community which currently has a "Best Current Practice RFC" (2488) and an "Ongoing Research RFC" (2760). Other possibilities include "programmer's guides," FAQs, and "How To" documents.

– Supporting the creation and distribution of intuitive cross-platform diagnostic utilities that allow gigabit networking users to determine if problems are within the local host platform, the remote host platform, and/or the network in between.

5 SHOWSTOPPERS

Packet loss was identified as one potential showstopper. Currently gigabit networking applications are primarily run in demonstration mode at prearranged times over minimally utilized networks. As these applications move more into the mainstream, they will start running on demand over shared networks, resulting in contention, congestion, and packet loss. Without some mechanism such as QoS to protect against packet loss, the applications will not be able to run at gigabit speeds over shared networks.

6 EVOLUTION CONSIDERATIONS

Many of the recommendations and processes described above would require a lifetime longer than that of the current NGI program to have a significant impact. As gigabit networking moves into the mainstream over the next several years, the packet loss issue will become much more critical. Security is another area that will be a definite post-NGI area of concern, as the applications involved in gigabit networking move from public demonstrations to production use of sensitive data over a shared network.

Measurement Roadmap

1 INTRODUCTION

The purpose of the Measurement Breakout Session was to identify current developments, challenges, and planned elements for gigabit networks during the next two years, and to partition recommendations into short-term best common practices and longer-term issues to direct or influence research programs and funding.

2 STATUS

Existing tools can serve for rudimentary assessment of end-to-end performance, topology, and workload, but capabilities are missing for segmentation, easy diagnosis, or application-specific characterization of problems. Two general categories of measurement tools support large-scale Internet data analysis systems: active and passive. Active tools inject packets into the network to ascertain performance or topology data. In testing performance, active measurements attempt to emulate the network performance an application would experience during the measured interval. These tools can affect and be affected by other (competing) traffic, so active measurements must be carefully designed and scheduled to minimize load on the network.

In contrast, passive measurement injects no traffic into the network. Instead, passive tools collect *(sniff)* traffic information as it passes by the measurement device, e.g., using promiscuous fiber tap. Alternatively, routers or switches themselves may embed statistic-gathering capabilities and provide access to those statistics via external queries (e.g., netflow, Simple Network Management Protocol). One can also use passive measurements to determine the actual performance and application experiences on a measured network, via timestamps of traffic entering and leaving the network. Although this last technique offers strong potential for Service Level Agreement (SLA) verification, no standard tools exist for doing so.

Key active and passive measurement tools in use (in no specific order):

Active Tools	Passive Tools
ping	cflowd
traceroute	coral/ocxmon
Skitter	PMA
Surveyor	flowscan
pathchar/pchar	netflow
TReno	NeTraMet
iperf/netperf	

For a more complete taxonomy, see *http://www.caida.org/tools/taxonomy.*

For a table of existing public measurement infrastructures, see *http://www.caida.org/analysis/performance/measinfra.*

3 CHALLENGES

There are general, technical, and programmatic challenges in deploying more broadly useful measurement capabilities.

3.1 General Issues

Challenge: evolve the field of Internet measurement from a set of disjoint, independent activities of network researchers, operators, and users toward an integrated service of the network that can support diagnosis of problems in real-time.

Obstacles include:

– The infrastructure is growing too fast to maintain a well-supported, comprehensive set of tools.

– We do not have good mechanisms or infrastructure to store, format, process, access, and visualize huge measurement data sets.

Measurement Roadmap cont.

– Little attention has been given to correlating data across many different sources or data type (topology, performance, workload, routing).

– Workload and performance metrics are them selves at fairly primitive states.

– Instrumentation can be hard to deploy at relevant locations: the “core” is splintered and competitive; the resulting limited access to data renders it difficult to argue that any given data set is “representative” (if users can get access to it at all).

– Negative perceptions regarding quality and legitimacy of data and methodology are driven by explosive growth.

3.2 Technical Issues

Higher bandwidths render most measurement challenges harder but not qualitatively different. One interesting historical aspect of Internet infrastructure evolution is the fact that the “advanced” higher bandwidth networks have typically been quite underutilized. Since performance problems often do not arise until networks are put under heavy load, measurement and diagnosis of such problems may ironically receive little research attention from the high-end community until congestion sets in. At that point it may be too late (or the focus may be on yet higher bandwidth rather than on engineering the current bandwidth better).

Note that bandwidth increases are not particularly relevant to many network performance aspects that have already required attention for years. (Higher bandwidth will just result in unsolved problems remaining unsolved.) Improving TCP performance, privacy (high performance and transparent encryption), standard metrics and formats for measurements, mechanisms for application profiling, and methods for pinpointing the location of network problems are all generally bandwidth independent. Performance asymmetry, where measurements may show acceptable performance in one direction but not in the other, has for several years suggested the value in using GPS receivers to do one-way delay measurements. Recently it has become clear that research is needed in alternative mechanisms of distributed one-way timing measurement, since installing GPS receivers at measurement locations is often not feasible. (NIST, AT&T and others are experimenting with stable rubidium oscillators and other synchronization techniques to overcome this problem.)

On the other hand, a jump in bandwidth does matter for other aspects of measurement:

– The difficulty of collection, filtering, and load generation at high speeds scales with the difficulty of building routing and switching devices for those speeds. Indeed, many of the same technologies are involved.

– Further analyzing high-speed traffic requires greater storage and compute cycles to accomplish meaningful aggregation, filtering, indexing, feature extraction, and visualization of large data sets.

– While small error rates (10^-4) may have little impact on LAN communication, such error rates can significantly degrade achievable throughput over a longer-delay high bandwidth WAN connection (particularly if small MTUs are used). Most LAN diagnosis and measurement tools will not reveal these characteristics.

– Throughput and available bandwidth tests are significantly more subject to “noise” in the measurements at higher speeds, and pulling the necessary signal out is correspondingly more difficult, e.g.,

Measurement Roadmap cont.

there exists no legitimate tool to assess the available bandwidth of a link above OC-3 (without consuming all the bandwidth, a rather self-defeating test).

– Measurement Information Base (MIB) variables wrap more quickly at higher speeds, especially when 32-bit counters are used.

However, perhaps more challenging than bandwidth increases is the rapid change in transport technologies, which have for years evolved faster than monitoring technologies. No single tool (or small set of fixed tools) will suffice for long; programmability at the software and hardware level is essential, as are consistent conventions for invoking tests and presenting results. A network measurement "service" framework should support this functionality, including methods for authentication and access control for the services.

3.3 Programmatic issues

Challenge: provide a *persistent measurement infrastructure* that scales with network size and bandwidth, utilizing *measurement systems that adapt to events and conditions*, e.g., taking more measurements during diagnosis periods and otherwise generating little or no background traffic. The infrastructure must include an efficient architecture for minimal collection, distribution, archiving and accessing of data, including methodology, tools, algorithms, and database structure to hold observations at a wide range of temporal and spatial granularity. These measurement systems should be an *integral part of the initial design of equipment and network architecture,* rather than be retroactively fitted into existing systems. Particularly critical factors are procuring strategic measurement locations and ensuring the security and privacy of actual IP addresses and other potentially sensitive data.

4 RECOMMENDATIONS

4.1 Short-term Recommendations

Objective: Improve measurement capabilities in today's infrastructure.

– Create a system that allows decomposition of what are now end-to-end performance measurements into measurement segments from end nodes to access providers (e.g., gigapops), and among access providers.

– Integrate and leverage the existing infrastructure that is already deployed. (See: *http://www.caida.org/analysis/performance/measinfra*.)

– Obtain volunteer machines at strategic sites (HPC campuses, gigapops) to host test end-points for use in path diagnosis. Establish consistent use of hostname aliases for testing platforms (e.g., iperf.ca.nlanr.net, or iperf.jpl.nasa.gov). These hosts should consist of well-tuned, well-connected platforms to best support segmented network path testing. Six to ten machines on each of the HPC backbones should be sufficient to characterize end-to-end capabilities on these networks, assuming the core is generally clear of errors.

– Determine a list of network administrators at each site that should have access to the measurement mesh and be responsible for training end users how to use it.

– Develop access control and privacy mechanisms that provide sufficient confidence that people would deploy a programmable device in their infrastructure and allow others to conduct experiments and extract data from it.

– Build a user interface to the above infrastructure that provides meaningful displays of collected information.

Measurement Roadmap cont.

– Develop active measurements that are less invasive and do not provoke providers to defensive behavior.

– Develop one-way performance measurement alternatives to GPS requirements.

4.2 Long-Term Recommendations

Longer term recommendations in support of a well-instrumented network, with seamless integration of data from a variety of sources, will require development of:

– New distributed measurement infrastructures that enable adaptive and programmable measurements and that can react to specific events with appropriately triggered measurements (e.g., performance events, faults, intrusions, etc.).

– An integrated measurement environment that allows experiments to be defined and programmed across multiple levels (including FPGAs, distributed procedural languages, and high-level inference/rule-based systems), and where results of measurement experiments are available for others to use. (Requires consistent data format conventions and tools.)

– Passive measurement infrastructure at HPC sites to support workload characterization such as at: *https://anala.caida.org/CoralReef/Demos/*.

– Methods of aggregating, mining, and visualizing the massive data sets in ways that are useful to multiple users (e.g., dynamic feature detection).

– Useful reporting to providers, site administrators and end-users in formats that support differentiated service contracts and facilitate routing and planning.

– Techniques for passive acquisition of performance data (latency, loss, jitter) that can reduce the perceived need for active probing of infrastructures.

– Analysis of the effects of sampling, multi-time scale analysis of data sets.

– Hybrid approaches combining (correlating) passive and active measurement data analysis in a single experiment.

– User-friendly integration with network utilities and control systems.

– Require, as part of project contracts/budgets, that network designs integrate measurement functionality into operational control and diagnostic loops for network infrastructures, taking into account advances in hardware speed and memory/bus limitations, emerging media (Gigabit Ethernet, DWDM), IP security, and the reluctance of ISPs to use and/or share measurement results.

– Promote the use and availability of such metrics and tools not only within research networks but also throughout the existing Internet where legitimate core backbone data from multiple providers offers more rigorous testing and verification.

5 CONCLUSIONS

The thousands of separate networks that make up today's Internet have little economic incentive to devote resources to developing consistent, universal measurement standards. In fact, they have incentives to avoid measuring traffic and performance data: 1) the lack of standardized testing and reporting methodologies may expose them to potentially inaccurate comparisons, and 2) exposing failures unilaterally, especially when competitors are not doing so, puts them at a marketing disadvantage.

Individually, no single segment of the fragmented Internet industry can address traffic engineering from the perspective of a larger Internet system.

Measurement Roadmap cont.

Collectively, with the encouragement of federal research sponsorship, it is possible to use a broadly based government, academic and industry research approach to address the macroscopic longer-term issues.

The speed of network infrastructure growth and introduction of new transport technologies means no single tool or small set of fixed tools will provide adequate measurements. Very small error rates can have a significant impact on the performance of high bandwidth long-distance networks; the measurement tools must be able to detect and report what used to be considered insignificant. An integrated network measurement service framework that is adaptable at the hardware and software level is essential to provide scaling and to accommodate ever-changing transports. This framework needs to be accepted and deployed by a majority of networks to adequately support full end-to-end performance measurement and troubleshooting.

Middleware Roadmap

1 INTRODUCTION

Middleware represents a broad and diverse set of tools to facilitate the use of basic network capabilities by applications and users. At the lowest level, middleware includes network layer tools, such as DNS, DHCP, multicast, and QoS management. Core middleware covers identifiers, directories, authentication and authorization; two critical and emergent technologies in this area are LDAP (Light Directory Access Protocol) and PKI (Public Key Infrastructure). Above core middleware sits a variety of niche middleware services. For scientific applications and gigabit networks, important "upper" middleware services include co-scheduling of networked resources, distributed network storage services, and generalized resource discovery.

The challenges in developing middleware are as much policy as technology. There are a few core technologies that need development and proof of scalability. Policy issues begin with the organizational formalization of security and access processes, and then correlating them with other entities, nationally and internationally, in both legal and operational frameworks. Some desirable functionalities such as delegation and revocation of trust present both technical and policy difficulties.

As a result of this complexity, most gigabit demonstrations to date have ignored middleware per se and developed point solutions to solve their particular needs. These approaches are often insecure, inflexible, and inconsistent. A true middleware fabric will be necessary to significantly increase the deployment of such applications, and to eventually enable commercial use.

A typical academic user working with a variety of agency facilities will need middleware services from several contexts. Personal identification and preferences must operate with institutional electronic security credentials and directory services, and these in turn must work with the broader research and educational community. Finally federal approaches must integrate with these other fabrics. When all this is constructed, then particular scientific environments such as computational grids, tele-immersion and data mining can be layered on top to complete the middleware for scalable and robust scientific research programs. The result must work broadly but be particularly sensitive to the requirements of gigabit networks.

2 STATUS

Three key technology building blocks for middleware to support gigabit to the desktop were identified as security, resource management, and performance tuning.

The breakout group found it difficult to identify middleware requirements that specifically address gigabit networking. However, because high-performance networking is enabling a new class of collaborative, distributed applications, the group focused on middleware requirements for this class of application, regardless of the transfer speed involved.

Grid technologies constitute an important area of gigabit networking applications, because of the major role of wide-area distributed computing in gigabit applications. There are already considerable efforts to develop packages of software services to support grid technologies.

2.1 Globus

The Globus project is developing the fundamental technology that is needed to build computational grids, execution environments that enable an application to integrate geographically distributed instruments, displays, and computational and

Middleware Roadmap cont.

information resources. The Globus toolkit is an integrated set of basic grid services: resource management, security, information infrastructure, communication, fault tolerance, and remote data access.

2.2 Grid Forum

The Grid Forum is an organization whose role is to document "best practices," develop implementation guidelines and standards with respect to grid technologies.

2.3 Legion

The Legion project is developing middleware to connect networks, workstations, supercomputers, and other computer resources together into a metasystem that provides the illusion of working on a single, virtual machine. Legion is based on a single unified object model to allow new applications to easily interact with all parts of the existing system.

2.4 Internet2 Middleware Initiative (I2-MI)

The I2-MI is working towards deployment of core middleware services at I2 universities. When applications provide identification, authentication, authorization, directories, and security, this leads to competing and incompatible standards. By promoting standardization and interoperability, middleware will make advanced network applications much easier to use.

3 ISSUES AND CHALLENGES

– Security is an especially critical issue, given the distributed nature of many gigabit applications.

- Security adversely impacts performance.
- Security may interfere with the seamless interface that scientists want to be presented with.
- Traversing long and complex trust chains to establish security among autonomous systems represents an extreme test of emergent PKI infrastructures.

– Performance tuning to support gigabit applications is difficult; packet loss is catastrophic.

– Broad deployment of middleware services is required, so that users can take these services for granted.

– Middleware services will not be used unless they are deployed in the gigabit testbeds, and these services will not be deployed until the users demand them. This poses a chicken/egg type of problem.

– Interoperability is critical.

– Middleware currently consists of ad hoc processes; support is needed for consistency and for development of APIs.

– Performance issues (e.g., see Security above).

– Coping with growth in dynamic range of computing and networking.

– Integration over a large number of autonomous environments.

Middleware Roadmap cont.

4 RECOMMENDATIONS

– Deploy middleware services in the testbeds, and encourage users to experiment with them.

– Continue ongoing efforts such as Globus and Legion.

– Encourage applications developers within the scientific community to replace point solutions with infrastructure as it emerges. Establish and offer training programs to application developers.

– Develop Network Weather Service as middleware.

– Incorporate network-sensing middleware into Globus.

– Develop standards for middleware.

– Develop middleware (such as adaptive QoS) to enable the network to be reconfigured in real-time to support gigabit applications. Such reconfiguration would reflect policy, performance, and security information.

Integration Roadmap

1 INTRODUCTION

Integration is a catch-all term intended to include any miscellaneous elements essential to the delivery of gigabit networking to the desktop that would not be covered in one of the other workshop sessions.

2 STATUS

Demonstrating a gigabit application requires considerable engineering effort, even when the network is lightly loaded. Individual developers tend to engineer their own solutions to make their application work.

3 ISSUES AND CHALLENGES

– Gigabit networking to the desktop tends to push beyond the tested limits of hardware and software. Undocumented pathological behaviors tend to emerge, causing problems that are difficult to diagnose.

– Performance tuning is critical in several areas, including error correction, MTU tuning, and TCP tuning.

– Interoperability is also key.

4 RECOMMENDATIONS

– Develop measurement tools and methodology to enable identification and isolation of performance problems. Establish performance baselines for each step in a compound path so that extraordinary circumstances can be readily identified.

– Develop an adaptive TCP stack.

– Promote creation of a common trouble ticket system.

– Develop software libraries to enable applications to interact with the operating system, middleware, and the network and manage performance parameters (such as buffers and time-outs) in a consistent fashion across the multiple layers of the protocol stack.

– Users of multiple autonomous systems are likely to have a variety of identifiers and attributes across the environments. Develop tools to automate the cross-walk of these characteristics so that values in one domain can be appropriately mapped onto other domains.

– Develop a toolkit for using gigabit testbeds, including standard diagnostic procedures and a “best practices” document.

– Train campus networking personnel to work with high-performance networks.

– Provide educational services for application developers, so that individual developers will not reinvent solutions.

8

Application Roadmaps

Teleseminars/Telemeetings

Models in Real Time

Huge Databases

Remote Instrumentation

Teleseminars/Telemeetings Roadmap

1 INTRODUCTION

Teleseminars and telemeetings will be a highly significant source of Next Generation Internet traffic and also will raise substantial service and performance issues for many years to come.

2 APPLICATION CHARACTERISTICS

Teleseminars are distinguished by requiring very high resolution and audio/video fidelity primarily in one direction, with a feedback channel for audience questions. Telemeetings are distinguished by requiring potentially all-to-all multicasting of many user sources simultaneously, with difficult session control and feedback issues that generate important research problems.

3 ISSUES AND CHALLENGES

Participants in this Gigabit Networking breakout session defined and discussed four performance/quality levels of teleseminars/telemeetings:

– “Extreme conferencing”

– Studio-quality HDTV-based conferencing

– Room-based conferencing

– “Cookbook” desktop conferencing

Two parameters were identified that cut across all four levels:

– Real time vs. archival: Real time is the primary focus for gigabit networking.

– Scheduled vs. persistent: Scheduled is the primary focus for the first and second levels (because they require a crew), and persistent (i.e., anyone/anytime) is the focus for the third and fourth levels.

The goal is to develop a conferencing infrastructure for all four levels that will persist after the NGI program ends in September 2002. In the top three levels, session participants identified specific gigabit prototyping demonstrations (listed below) that ideally will be carried out by the LSN agencies and Internet2 universities during the next two years.

4 RECOMMENDATIONS

4.1 Extreme Conferencing

“Extreme Conferencing” was coined as an informal term to use to describe the maximum conferencing capability that is achievable. Whereas the current MBone technology works for at most a few dozen participant images, the DARPA Virtual Amphitheater project seeks to bring 200 to 1000 interactive participant images into a single amphitheater image of moderate quality and resolution, with a central stage of very high quality and resolution for panels and performances.

Each of the participant images would be generated with encodings delivering a few hundred kbps (say 60 x 100 pixels). The stage or panel images could be made up of several sources, each encoded with multi-megabit-per-second resolution and quality. The focus of this project is the digital processing and creation of the composite images in a video merge server, as well as developing the end-system architecture and researching the control issues. Virtual Amphitheater plans to put on demonstrations by September 2001 and deliver a complete prototype by September 2002.

4.2 Studio-quality HDTV-based Conferencing

Applications such as teleseminars on fine arts or biology, or remote consultation for radiology or surgery, truly require the full resolution and studio quality of High Definition Television (HDTV). The standard HDTV signal rates are 1.5 Gbps uncompressed, 200 Mbps studio quality, 40 Mbps contribution quality (used for feeds from television station affiliates to their networks), and 19.2 Mbps broadcast quality.

Teleseminars/Telemeetings Roadmap cont.

Gigabit IP network engineers need to learn how to routinely handle scheduled transmissions of HDTV. This will involve issues such as bandwidth, traffic engineering, quality, measurement, multicast, and control. In addition, gigabit networking researchers must explore several difficult latency and quality issues related to HDTV-based teleseminars and telemeetings, such as forward error correction (FEC) vs. buffering and retransmission, user quality perception, error rates, and delay introduced by compression. The specific application environment also makes a big difference, e.g., whether the studio-quality HDTV-based conferencing is being used to support a live teleseminar, or remotely access an archived teleseminar, or perform a real-time remote scientific experiment.

The following specific studio-quality HDTV-based conferencing demonstration possibilities were identified:

– University of Washington research seminars (David Richardson: drr@washington.edu) *(Note: The university of Washington is not currently funded adequately to continue their current level of leadership in this area, so this could prove to be a showstopper.)*

– Distributed studio (Sony)

– NRL HDTV experiments (Hank Dardy)

– NASA: “DreamTime” initiative for International Space Station (Sony), museums (e.g., American Museum of Natural History)

– NIH/VA remote consultation prototyping

– NOAA National Hurricane Center (Bill Turnbull)

4.3 Room-Based Conferencing

Room-based conferencing is available now and used at Berkeley MASH workshops (www.openmash.org). The focus here is to enable a conference to be supported by at most one lightly trained operator. For gigabit networking, existing room-based conferencing systems need to be upgraded to much higher quality and resolution than is typical today. Research continues to be needed on such important issues as floor control, audio leveling, echo cancellation, and audio/video integration.

The following specific room-based conferencing gigabit networking demonstration possibilities were identified:

– Access Grid video chat rooms

(http://www-fp.mcs.anl.gov/fl/accessgrid/default.htm)

– NASA HDTV videoconferencing prototyping

– MPEG2 videoconferencing multicast

– Internet2/University of Akron interoperability between Litton and Minerva systems

– Cisco IPTV

4.4 “Cookbook” Desktop Conferencing

The focus here is to support spontaneous multi-desktop conferences with no operators required. Examples are VIDE (www.vide.net) and VRVS (wwwvc.cern.ch).

It is not clear whether this area could see a true gigabit networking demonstration during the next two years. One possibility was to begin using digital video and Firewire-based videoconferencing systems.

Models in Real-Time Roadmap

1 INTRODUCTION

1.1 Purpose

The purpose of this session was to prepare roadmaps for the development of four to five candidate "Models in Real Time" application demonstrations to be completed during the NGI program. These application demonstrations using computational models would require transfer rates exceeding 500 Mbps and require the use of Gbps networking technologies. These application demonstrations would contribute to the application legacy of the NGI program. The status and issues of applications and the processes and requirements for application demonstrations would be documented.

1.2 Definition

The group first defined the concept of "Models in Real Time." In contrast with other definitions, "real time" in this context refers to creating a model of a process that is understandable in human time. The applications in this category involve a human somewhere in the loop, and the visualization of, or interaction with, an event in human time is considered real-time—real-time being a relative term, dependent on the context in which it is used.

1.3 Synergy with Huge Databases

At the beginning of the session it was suggested that the two sessions, Huge Databases and Models in Real Time, should be integrated into one because models in real time require the use of huge databases. It was decided to pursue a separate discussion and then later, meet with the Huge Databases group. In the discussion that followed regarding the synergy between the two areas, several types of applications were identified that demonstrate this synergy:

– Database multicasting issues

– Real-time models with interactive steering

– Visualization of computations

2 STATUS

The group discussed the rationale for the workshop's focus on demonstrations rather than on long-term solutions. The point was made that this is the third year of the government-sponsored NGI initiative, for which Congress mandated performance goals. The goals state that 10 applications requiring Gbps connectivity rates be demonstrated. The ability to meet these goals was questioned, as it is now halfway through the NGI period. These applications should have been identified and funded during the first year of the NGI, with demonstrations following. It will be possible to finish applications that have already been started before the end of the two years, but projects not yet started will not finish in time to meet the deadline. However, agencies that were given money should influence the next round of funding in order to get the job done and demonstrate these applications.

3 POTENTIAL APPLICATIONS

Candidate applications of the organizations represented in the discussion were described. However, since the representatives were not those people actually working on any of the applications, there were no in-depth discussions of requirements. The applications discussed are as follows.

3.1 Potential DoD Applications

– Battlefield simulations with tens of thousands to hundreds of thousands of elements interacting

3.2 Potential DoE Applications

– Combustion Corridor (will be restarted, but a new version is unlikely to be ready for demonstration by September 2002)

Models in Real-Time Roadmap cont.

3.2 Potential DoE Applications cont.

– High energy physics: particle data grid
  - model is not in real time
  - remote instrumentation is the same as or similar to data collection in real time
  - compare data from remote instrumentation with data created by the computational model (where computational data is created either before or concurrently with data collection)

– Climate data modeling

3.3 Potential NASA Applications

– Partial or complete aircraft models
  - distributed
  - multi-disciplinary

– Propulsion models
  - object oriented
  - distributed
  - multi-disciplinary

– Comparison of wind tunnel results to computational results—physical model vs. mathematical model

3.4 Potential NOAA Applications

– Immersive technologies in collaborative mode

– Ocean circulation model
  - interactive fisheries and oceanography
  - distributed access

– Weather models
  - Real-time radar data into forecast models

The basic models of many of the above applications are already operating, but adding a real-time component requires significant change. These changes will not occur in the NGI timeframe. It is important to note that NGI will be the enabler for making these applications possible.

4 ISSUES AND CHALLENGES

The group discussed some of the problems and issues involved in meeting the two-year time limitation remaining on the government's five-year NGI initiative.

4.1 Collaboration Between Applications Personnel and Network Personnel

As an example, NOAA is just getting started on developing applications to use gigabit networks in the wide area. NOAA collaborations are also just beginning between applications personnel and network personnel.

Modelers are busy moving from individual sequential applications to parallel applications and thus do not think about the networks involved in running their applications. This is possibly because they have not yet expanded their applications to the point where networking is required. Since networks have not been used much in modeling, NGI should focus on merging applications and networks.

4.2 Modeling Difficulty

It is harder to get large-scale models going because it is such a big effort, and smaller distributed access, such as a group of people using the Internet for collaborative research, has not worked for these types of applications.

4.3 Funding

A big issue affecting the ability to perform applications demonstrations is funding. For example, DoE's participation in NGI was terminated, so applications like the Combustion Corridor demonstrated at this workshop have been discontinued. To be called "NGI", applications must come at least in part from NGI-funded projects, and money is always an

Models in Real-Time Roadmap cont.

issue, especially as networking and applications have often been funded separately. Partnership with other institutions on projects is a good thing and should be encouraged.

4.4 Questions

Some questions require answers, such as:

– Do applications have to be in a particular area or areas?

– Can DoE form partnerships with others? Could DoE perhaps demo if political considerations can be overcome?

– Can funding be provided if a new application is identified?

5 RECOMMENDATIONS

– Bring together separate network and application groups.

– Avoid decoupling network and application development.

– Network development should stress latency reduction in addition to raw bandwidth.

– Follow-up discussions with applications developers should be held to provide the required information.

Huge Database Roadmap

1 INTRODUCTION

Note: *Participants of this breakout session felt that "large databases" was more appropriate terminology than "huge databases." However, the workshop organizers decided to retain the original terminology to reflect the focus on gigabit networking.*

Gigabit networks will enable new data-intensive applications involving remote and distributed data. Up to now, data-intensive applications have generally used local data and local computational resources. Data-intensive computations require a special data management infrastructure to compute gigabyte-size chunks obtained from querying terabyte-size data sets extracted from petabyte-size archives.

Extending this data management infrastructure to work with remote and distributed data is a fundamental scientific challenge. In particular, the data management infrastructure must be carefully balanced with the computing infrastructure as well as the networking infrastructure. The challenge here is not only to reach across a network and pull a previously identified megabyte of indexed data from a terabyte-size database, but also to compute with gigabyte-size chunks of data buried in terabyte databases and petabyte archives which are distributed over the network and to discover new patterns and correlations along the way.

2 STATUS

We imagine that early adopters of gigabit applications will share some of the following characteristics: They will be well-equipped research scientists connected to data sources and computational facilities via wide area OC-12 or OC-48 and local campus gigabit Ethernet. They will have terabytes of local storage for caching and intensive analysis. They will have local computational resources consisting of 32- or 64-node workstation clusters and software supporting different styles of interaction, such as interactive exploration, coffee-break computations and overnight computations.

There will be a number of sources of data-intensive gigabit applications, including:

– *Huge scientific databases*. By 2002, there should be a few dozen petabyte archives with the appropriate data extraction mechanisms required to support gigabit applications. Generally this data is created by scientific instruments and platforms. Examples include EOS, astronomical databases, high energy physics databases, nuclear physics databases, climatic databases, and databases of network data. Increasingly these platforms are themselves distributed.

– *Simulations*. With the increasing power of supercomputers and computer clusters, simulations can today easily produce terabyte-size data sets. Another source of gigabit applications arises when remote scientists try to access these simulations.

– *Multimedia data*. Multimedia data, including video and HDTV, are becoming increasingly common not only for entertainment but also for remote learning and oral history.

3 CHALLENGES

Data-intensive gigabit applications face special challenges, including:

– *Exploiting network bandwidth without tuning.* Currently exploiting network bandwidth requires a lot of hand tuning. Unless simpler mechanisms are developed, gigabit applications for working scientists, and not just short-term demonstrations, will not be possible.

Huge Database Roadmap cont.

– *End-to-end requirements matching*. The end-to-end requirements for gigabit applications are demanding. Data must be extracted off of disk and tape, processed, moved across the network, processed at the other end, visualized, and stored onto disk and tape. Extracting and storing data presents additional requirements that are not present with computational and streaming data applications.

– *Big data requires big CPUs*. There are two modes of interacting with huge data sets: extracting small subsets for further analysis and studying and looking for patterns and correlations in huge sets. The latter (data analysis or data mining) requires huge computational resources.

– *Geographic distribution*. As the amount of data grows and as the number of people interested in accessing it grows, new problems emerge, including data distribution policies, data caching policies, data multicasting, etc.

– *Better network instrumentation*. Current network instrumentation does not directly support the needs of data-intensive gigabit application developers. For data-intensive gigabit applications to be developed, better tools for measuring data movement and network utilization are needed.

– *The familiar problems*. Just because the amount of data is growing and the available bandwidth is growing does not mean the old “hard” problems have been solved. For example, it is not obvious that from a practical viewpoint our ability to manage metadata has gotten any better during the past decade. Also, scaling data management, feature extraction, and data mining to terabyte-size datasets are still basically open problems. Finally, huge datasets and gigabit applications require that the basic tenets of hierarchical storage be built into the system from the ground up. This is still not the way most systems are designed today.

4 RECOMMENDATIONS

– Create a short list of potential gigabit demonstrations to consider for inclusion:

- Real time instruments generating huge data streams
- Bioinformatics/molecular databases
- Real time weather prediction
- Real exploitation of network data
- High energy physics databases

– Create a timeline for rollout of gigabit demonstrations (12-27 months).

– Provide adequate support for software development and demonstrations.

– Provide incremental funding to a short list of existing projects to add gigabit applications as a deliverable.

5 EVOLUTION CONSIDERATIONS

Data-intensive gigabit applications require that there be no bottlenecks in the end-to-end path that data takes from remote disk to local application and that this be achieved with extensive network tuning. With a typical data-intensive application spreading data across hundreds of disks and tapes, building a data-intensive gigabit application will place special requirements on the software and network infrastructure being developed.

– Network Infrastructure

- Local gigabit networks and wide area OC-12 and OC-48
- End to end network performance out of the box

Huge Database Roadmap cont.

  - Easier end to end high bandwidth capability
  - Effective middleware
  - Improving latency issues in underlying networks
  - Good specification of application network requirements (latency vs. bandwidth)
- Software Infrastructure
  - Middleware and other software for working with huge amounts of remote and distributed data
  - Aligning the application to the middleware and to the network
  - Appropriate data extraction mechanisms from petabyte archives
  - Appropriate query mechanisms for terabyte-size databases
  - Data multicasting
  - Metadata support for gigabit data-intensive applications
  - Resource location management for gigabit data-intensive applications
- Testbeds
  - Interim testbed to provide a better understanding of I/O vs CPU issues for remote data analysis, distributed data mining and wide area interactive exploration of data
- Funding
  - Consistent funding of petabyte data centers
  - Developing human resources to make this happen

Remote Instrumentation Roadmap

1 INTRODUCTION

Remote instrumentation refers to the use of scientific instruments operated remotely over a network. Examples are astronomical telescopes on high mountain tops in Hawaii or Chile that are remotely operated by university researchers at their home sites, or one-of-a-kind advanced radiation sources operated by DOE at national laboratories and used remotely by DOE-supported research scientists. Remote instrumentation raises a host of difficult technical issues such as low latency for support of interactive applications, multicast for support of group communications, and authorization and authentication security for restricting access to authorized users only.

This Remote Instrumentation breakout session discussed the particular and unique requirements involved in supporting remote instrumentation applications, and identified several remote instrumentation applications that could be used to demonstrate gigabit networking within the upcoming two-year timeframe.

2 APPLICATION REQUIREMENTS

Several particular needs were identified for remote instrumentation. Some of these are unique to remote instrumentation, others apply to other application areas as well. The list of needs identified is as follows:

- Authorization and authentication of remote users
- Scheduling and reservation of experiment time
- Encryption
- Quality of Service
- Low latency routing algorithms (shortest and fastest possible route)
- Scalability for multiple users
- Different tolerance levels and requirements (willingness to accept some errors or latency)
- Visualize or create feel for latency and adapt to it (compensate via agent)
- Multiple observers, passive and active (users can “pass-off” or request control)
- Heterogeneous video streams for passive observers
- Use of multicast vs. unicast
- Fixed locations vs. mobile

3 ISSUES AND CHALLENGES

3.1 Telemicroscopy

Telemicroscopy is the remote use of unique microscope facilities. Users need near real-time interaction with the microscope to do such tasks as change magnification level, enhance focus, change brightness, and move the specimen. This is accomplished through remote user control and live streaming video of the specimen.

The use of gigabit networks would allow real-time video presented in various formats such as Motion JPEG and digital video. With increased bandwidth, users could receive clearer and higher resolution images, allowing for more accurate analysis of the specimen. Higher data transfer rates would enable users to adjust focus and see changes almost immediately. This would greatly shorten the current 30- to 40-minute process of locating specific areas on a specimen through scanning and searching the small viewing area at low magnification.

Lower latency would allow users to react quickly as if local, e.g., when trying to catch which video frame shows an adjustment made in real time. Current auto-focusing algorithms are geared toward maintaining a stable optical image while gross changes, such as when switching viewing areas, require

Remote Instrumentation Roadmap cont.

manual intervention. Typically 60 to 100 detailed images of a specimen are taken, requiring the user to tilt the specimen at different angles and adjust focus in real time.

Another problem is network bottlenecks, which can occur, for example, when the UDP stream for digital video hogs most of the router buffer space and does not allow commands sent via TCP to reach the destination. All these problems are amenable to amelioration through gigabit networking research.

One such microscopy center links the San Diego Supercomputer Center (SDSC) with Osaka University in Japan, which houses the world's largest transmission electron microscope. SDSC is able to remotely use the Osaka microscope via high-speed networks. From the data that is collected through user interaction with the microscope and the specimen, a 3D model is constructed by SDSC supercomputers for further analysis after the real-time observation period ends.

3.2 Remote Telepresence Surgery System

Remote telepresence allows a surgeon working with visual 3D images to perform surgery remotely via the use of a robot. Practical applications would be for use in the battlefield as well as in microscopic surgery. The robot follows the surgeon's hand movements and can ameliorate some human limitations such as dexterity and tremor. Remote surgery would be useful if a specialized surgeon is required but not available at the remote location.

Remote telepresence can also be used for training and simulation purposes, reducing training cost and learning curve. Remote telepresence has already been approved for abdominal surgery and could be used for cardiac surgery.

One big problem with remote surgery is the latency introduced through the network. The remote surgical instruments have to be positioned very accurately; tiny differences in time between issuing a command and the machine receiving it could end up with catastrophic results. Haptic ("touch") feedback to the surgeon requires less than a few milliseconds of delay to be useful. Assured quality of service from the network is essential for this application area, as well as the use of specialized assured (low) latency routing algorithms.

3.3 Remote Telescopes

With telescope equipment becoming fully digital, the demand for remote control of telescopes is increasing. Typically today, observers have to travel to the telescopes, which are frequently located in remote, high altitude areas. Researchers are given reserved windows, or shares of time, to use the telescope. Remote control would save considerable expensive travel cost and allow better use of shared observing time. Since the coordinates of the area to be observed are known ahead of time, there is no need for scanning and searching as required with microscopes. The data-taking requirements of multiple observers can frequently be interleaved with each other, increasing scientific efficiency.

Requirements for remote telescopes include large amounts of bandwidth for quick bursts of initial observation data from the telescope to the user, and corrective commands from the user to the telescope. Once the telescope parameters have been properly adjusted, the telescope observation frequently takes many minutes or in some cases an hour or more of time, during which no data or commands are flowing. Once the entire data observation has been completed, then the re-

Remote Instrumentation Roadmap cont.

searcher is interested in receiving the data immediately so that he or she can begin analyzing the science content. Frequently one observation will lead to another a few days later. With on-site observation, researchers will not have processed the observations in detail before leaving the telescope site, whereas with remote telescope operation, researchers can make efficient multiple observations even when these observations are separated by days or weeks.

3.4 Other Possible Remote Instrumentation Applications

Other applications were identified that could lead to gigabit networking applications in the area of remote instrumentation. These include:

– Neptune Project. Underwater fiber plant base around tectonic plates for making measurements of plate shifts.

– Shake Tables. Simulate earthquakes remotely. Has commercial usage.

– Wind Tunnels. Retrieve data from wind tunnel experiments remotely (Boeing).

– Integrate Weather Stations and Large Meteorological Databases.

– Hazardous Environments. All work has to be done by remotely controlled robots.

– Matisse. See *www.cnri.reston.va.us/matisse.*

– X-ray Crystallography and other applications of DOE advanced radiation sources.

– High Energy Nuclear Physics (DOE and CERN).

4 SHOW STOPPERS

– Shared LANs: The real-time nature of remote instrument observations demands deterministic access along the entire end-to-end network path. Network segments that use shared media access protocols produce unpredictable access patterns that can make the application difficult or impossible to operate.

– Latency-specific routing: The interactive nature of remote instrument observations demands low or predictable latency along the entire end-to-end network path. Current Internet routing algorithms are designed to forward packets based on the "best" path, which may not be path with the lowest or most predictable latency. New algorithms need to be developed to correct this problem.

– Where is money going to come from 2+ years from now?

– Who will use it in 2 years? Will it be practical and marketable?

Perspectives

9

Current Status and Future of Gigabit Networking

MOHAMMED ATIQUZZAMAN, UNIVERSITY OF DAYTON, OHIO

PRESENT STATE OF THE NETWORK

TCP/IP is the internetworking technology that glues together all the computers in the Internet. Because TCP/IP was originally designed for applications such as remote login and file transfer, the current TCP/IP does not provide Quality of Service guarantees to real time applications that may be required in gigabit networks.

CURRENT EFFORTS

A number of efforts are currently underway to incorporate QoS in future gigabit networks. The efforts are briefly described below.

Integrated Services: Integrated Services (IS) uses RSVP signaling protocol to reserve resources in the network for every connection (flow) that has to be set up. RSVP signaling is used to pass the flow parameters to the Network Elements (typically routers) for admission control purposes. The Network Elements determine whether enough resources are available, and then signal the hosts whether the flow can be setup. Implementing admission control in addition to policing of the individual connections provides QoS guarantees to individual applications. Storing the states of many individual connections gives rise to scalability problems at the routers in the core network.

Differentiated Services: The scalability problem of Integrated Services resulted in a second effort called the *Differentiated Service* (DS). The DS is based on classifying the packets (depending on their service requirements) into a number of classes at the ingress to the network. Within the core network, all the connections belonging to a particular class are combined into one group and receive aggregate behavior. DS does not guarantee QoS for individual applications.

Asynchronous Transfer Mode (ATM): Asynchronous Transfer Mode (ATM) has been designed from the onset with QoS built in the network. Currently, ATM is being deployed in the backbone of many gigabit networks with TCP/IP being used at the edge of the network (such as desktops and LANs).

NEXT GENERATION GIGABIT NETWORK

A possible architecture for Next Generation Gigabit networks is to use Integrated Services in the edge network with Differentiated Services and/or ATM at the core network, with some part of the network being run over satellite links having long propagation delays. Connecting different networks, however, gives rise to a range of issues relating to mapping of services and service parameters, protocol conversion, interoperability, end to end QoS and performance.

Perspectives

RONN RITKE, NATIONAL LABORATORY FOR APPLIED NETWORKING RESEARCH

The backbone network speeds have increased from T3 (NSFnet) to OC3, OC12 (vBNS) and then to OC48 rates (Abilene). As backbone speeds have increased, some of this speed has reached closer to the end user. We now have Gigabit Ethernet and computers (Apple Power Mac G4) with built-in Gigabit Ethernet capabilities that potentially offer higher bandwidth to the end user. But for most people, the last-mile problem remains. How to detect and deal with bottlenecks that reduce the bandwidth is a major problem. While high-bandwidth is very useful and has allowed for the development of high-bandwidth and/or low-latency applications, network design should take into account other factors.

Two questions can begin the dialog:

1. Is speed the only facet to examine?

2. Should other metrics be used?

Appendices

10

POCs, Sponsors

Attendees

Glossary

Points of Contact

WORKSHOP HOST

Ken Freeman
NASA/NREN
650.604.1263
kfreeman@arc.nasa.gov

CO-LEADS OF THE ORGANIZING COMMITTEE

Kevin Jones
NASA/NREN
650.604.2006
kjones@arc.nasa.gov

Richard desJardins
NASA/NREN
650.604.4764
rdesjardins@arc.nasa.gov

Mark Foster
NASA/NREN
650.604.1809
mafoster@arc.nasa.gov

Marjory Johnson
NASA/NREN
650.604.6922
mjjohnson@arc.nasa.gov

KEYNOTE

Raj Jain
Ohio State University
Jain@CIS.Ohio-State.Edu

DEMONSTRATIONS

Land Speed Record/Internet2

Terry Gibbons
USC Information Sciences Institute
Tgibbons@isi.edu

Combustion Corridor—Visapult

Wesley Bethel
Lawrence Berkeley National Laboratory
Ewbethel@lbl.gov

Brian Tierney
Lawrence Berkeley National Laboratory
bltierney@lbl.gov

Robert F. Lucas
Lawrence Berkeley National Laboratory
rflucas@lbl.gov

Project DataSpace

Robert Grossman
University of Illinois at Chicago
grossman@uic.edu

Marco Mazzucco
University of Illinois at Chicago
marco@newton.math.uic.edu

Emory Creel
University of Illinois at Chicago
emory@lac.uic.edu

Digital Sky Virtual Observatory

Joseph Jacob
Jet Propulsion Laboratory
Joseph.jacob@jpl.nasa.gov

David Curkendall
Jet Propulsion Laboratory
Dwc@jpl.nasa.gov

Virtual MechanoSynthesis (VMS)

Jon Guice
NASA Ames Research Center
jon@ptolemy.arc.nasa.gov
650.604.2822

Chris Henze
NASA Ames Research Center
Henze@arc.nasa.gov
650.604.3959

Points of Contact

DEMONSTRATIONS CONT.

Visible Human

Brian Athey
University of Michigan
Bleu@umich.edu

Thomas Hacker
University of Michigan
Hacker@umich.edu

High Definition TV

David Richardson
University of Washington
Drr@washington.edu

PANELISTS

Richard Carlson
Department of Energy
Carlson@er.doe.gov

Phillip Dykstra
WareOnEarth Communications, Inc.
Phil@wareonearth.com

Ian Foster
Argonne National Laboratory/
University of Chicago
foster@mcs.anl.gov, www.mcs.anl.gov/~foster

Kay Howell
National Coordination Office for Computing, Information and Communications
Howell@ccic.gov

Wesley K. Kaplow,
Qwest Government Systems
Wesley.kaplow@qwest.com

John Wroclawski
Massachusetts Institute of Technology
jtw@lcs.mit.edu

TECHNOLOGY PRESENTATIONS

Developing the Infrastructure

Jim Gimlett
Network Elements
Jgimlett@networkelements.com

Basil Irwin
University Corporation for
Atmospheric Research
Irwin@ucar.edu

Leonid Kazovsky
Stanford University
Leonid@ocrl.stanford.edu

Matthew Mathis
National Laboratory for Applied
Networking Research
Mathis@psc.edu

Nick McKeown
Stanford University
Nickm@stanford.edu

Tools for Gigabit Networking

Steve Corbato
University Corporation for Advanced
Internet Development
Corbato@internet2.edu

Cas D'Angelo
Georgia Institute of Technology
Cas.DAngelo@OIT.GaTech.EDU

Joseph B. Evans
University of Kansas, Lawrence
evans@ittc.ukans.edu

William Lennon
Lawrence Livermore National
Laboratory
Wjlennon@llnl.gov

Michael O'Connor
Brookhaven National Laboratory
moc@bnl.gov

Points of Contact

Jerry Sobieski
Mid Atlantic Crossroads
jerrys@maxgigapop.net

TESTBED REPRESENTATIVES

Internet2

Ted Hanss
Ted@internet2.edu

Paul Love
Epl@internet2.edu

ADDITIONAL NGI AGENCY REPRESENTATIVES

DARPA

Mari Maeda
Mmaeda@darpa.gov

DOE

Richard Carlson
Carlson@er.doe.gov

Thomas Ndousse-Fetter
Tndousse@er.doe.gov

NCO

Grant Miller
Miller@ccic.gov

Sally Howe
Howe@ccic.gov

NIH

Jules Aronson
Aronson@nlm.nih.gov

Michael Ackerman
Ackerman@nes.nlm.nih.gov

Michael Gill
Mike_gill@nlm.nih.gov

NIST

Doug Montgomery
Dougm@nist.gov

NOAA

William Turnbull
Wturnbull@hpcc.noaa.gov

NSA

Bill Semancik
Wjseman@lts.ncsc.mil

NSF

George Strawn
Gstrawn@nsf.gov

Aubrey Bush
Abush@nsf.gov

Karen Sollins
Ksollins@nsf.gov

LOGISTICS

Sally Miller
NASA/NREN
650.604.5411
smmiller@arc.nasa.gov

Pat Kaspar
NASA/NREN
650.604.5391
pkaspar@arc.nasa.gov

SPONSORS

Qwest Communications
http://www.qwest.com/

Cisco Systems
http://www.cisco.com

Attendees

Michael Ackerman
National Library of Medicine
ackerman@nlm.nih.gov

Jules Aronson
National Library of Medicine
aronson@nlm.nih.gov

Brian Athey
University of Michigan
Bleu@umich.edu

Mohammed Atiquzzaman
University of Dayton
atiq@udayton.edu

Raman Azizian
NASA Marshall Space Flight Center
raman.azizian@msfc.nasa.gov

Wesley Bethel
Lawrence Berkeley National Laboratory
ewbethel@lbl.gov

Shoshana Billik
NASA Ames Research Center
shoshana@nas.nasa.gov

Nichole Boscia
NASA Ames Research Center
nboscia@nas.nasa.gov

George Brandt
Sandia National Laboratory
brandt@ca.sandia.gov

Richard Carlson
Department of Energy
carlson@er.doe.gov

Helen Chen
Sandia National Laboratory
hycsw@ca.sandia.gov

Kim Claffy
CAIDA
kc@sdsc.edu

Steve Corbato
Internet2
corbato@internet2.edu

Emory Creel
University of Illinois, Chicago
emory@lac.uic.edu

David Curkendall
Jet Propulsion Laboratory
dwc@jpl.nasa.gov

Cas D'Angelo
Georgia Institute of Technology
Cas.dangelo@oit.gatech.edu

Francesca Davis
NASA Marshall Space Flight Center
Francesca.Davis@msfc.nasa.gov

Claudia deLuna
Jet Propulsion Laboratory
claudia.de.luna@jpl.nasa.gov

Richard desJardins
NASA Ames Research Center
rdesjardins@arc.nasa.gov

Dan Duffy
NASA Goddard Space Flight Center
dan.duffy@gsfc.nasa.gov

Larry Dunn
Cisco Systems
ldunn@cisco.com

Phillip Dykstra
WareOnEarth Communications
phil@wareonearth.com

Joseph Evans
University of Kansas
evans@ittc.ukans.edu

Alan Federman
Raytheon
afederman@mail.arc.nasa.gov

William Feiereisen
NASA Ames Research Center
wfeiereisen@mail.arc.nasa.gov

James Finch
Wang Government Services
jim.finch@csoconline.com

Attendees cont.

William Fink
NASA Goddard Space Flight Center
bill@wizard.gsfc.nasa.gov

Dave Foltz
NASA Glenn Research Center
dfoltz@grc.nasa.gov

Ron Ford
Wang Government Services
ron.ford@csoconline.com

Ian Foster
Argonne National Laboratory
foster@mcs.anl.gov

Mark Foster
NASA Ames Research Center
mafoster@mail.arc.nasa.gov

Ken Freeman
NASA Ames Research Center
kfreeman@mail.arc.nasa.gov

Pablo Garcia
SRI International
pgarcia@unix.sri.com

Andy Germain
NASA Goddard Space Flight Center
andy.germain@gsfc.nasa.gov

Terry Gibbons
USC Information Sciences Institute
tgibbons@isi.edu

Mike Gill
National Library of Medicine
mike_gill@nlm.nih.gov

Ray Gilstrap
NASA Ames Research Center
rgilstrap@arc.nasa.gov

Jim Gimlett
Network Elements
jgimlett@networkelements.com

Howard Gordon
National Security Agency
flash@super.org

Terry Gray
University of Washington
gray@washington.edu

Bryan Green
NASA Ames Research Center
bgreen@nas.nasa.gov

Robert Grossman
University of Illinois, Chicago
grossman@uic.edu

Jon Guice
NASA Ames Research Center
jon@ptolemy.arc.nasa.gov

Thomas Hacker
University of Michigan
hacker@umich.edu

Martin Hadida
University of California - San Diego/ San Diego Supercomputing Center
marty@sdsc.edu

David Hajazin
NASA Marshall Space Flight Center
david.hajazin@msfc.nasa.gov

Ted Hanss
Internet2
ted@internet2.edu

Fritz Hasler
NASA Goddard Space Flight Center
hasler@gsfc.nasa.gov

Chris Henze
NASA Ames Research Center
chenze@nas.nasa.gov

Matthew Hodge
University of Washington
mihodge@cac.washington.edu

Kay Howell
National Coordination Office
howe@ccic.gov

Basil Irwin
University Corporation for Atmospheric Research
irwin@ucar.edu

Attendees cont.

Joseph Jacob
Jet Propulsion Laboratory
Joseph.Jacob@jpl.nasa.gov

Raj Jain
Nayna Networks
jain@acm.org

Marjory Johnson
NASA Ames Research Center
mjjohnson@riacs.edu

William Johnston
NASA Ames Research Center
wej@nas.nasa.gov

Kevin Jones
NASA Ames Research Center
kjones@mail.arc.nasa.gov

Wesley Kaplow
Qwest Communications
wesley.kaplow@qwest.com

Pat Kaspar
NASA Ames Research Center
pkaspar@mail.arc.nasa.gov

Leonid Kazovsky
Stanford University
leonid@ocrl.stanford.edu

Richard Kitterman
Raytheon
rkitterman@mail.arc.nasa.gov

Ken Klingenstein
Internet2
Klingenstein@internet2.edu

Stephen Klotz
NASA Ames Research Center
klotz@nas.nasa.gov

Gordon Knoble
NASA Goddard Space Flight Center
Gordon.Knoble@gsfc.nasa.gov

Hugh LaMaster
NASA Ames Research Center
lamaster@nas.nasa.gov

Foo Lee
NASA Ames Research Center
flee@mail-aix.arc.nasa.gov

Will Leland
Telcordia Technologies
wel@research.telcordia.com

William Lennon
Lawrence Livermore National Laboratory
wjlennon@llnl.gov

Joseph Loiacono
NASA Goddard Space Flight Center
jloiacono@nastg.gsfc.nasa.gov

Isaac Lopez
NASA Glenn Research Center
ilopez@grc.nasa.gov

Paul Love
Internet2
epl@internet2.edu

Dan Magorian
University of Maryland
magorian@nts.umd.edu

Michael Manyin
NASA Goddard Space Flight Center
manyin@agnes.gsfc.nasa.gov

Matthew Mathis
National Laboratory for Applied Network Research/
Pittsburgh Supercomputing Center
mathis@psc.edu

Marco Mazzucco
University of Illinois, Chicago
marco@dmg.org

Nick McKeown
Stanford University
nickm@stanford.edu

Nancy McKown
NASA Ames Research Center
nmckown@arc.nasa.gov

David Meyers
NASA Ames Research Center
dmeyers@arc.nasa.gov

Attendees cont.

Grant Miller
National Coordination Office
miller@ccic.gov

Doug Montgomery
National Institute for Standards and Technology
dougm@nist.gov

Craig Moyers
Raytheon
cmoyers@mail.arc.nasa.gov

Klara Nahrstedt
University of Illinois
klara@uiuc.edu

Vishy Narayan
Coreon Inc.
vishy@ieee.org

Thomas Ndousse
Department of Energy
tndousse@er.doe.gov

Dwayne Nelson
IBM
enelson@almaden.ibm.com

JoAnn Nelson
NASA Ames Research Center
janelson@arc.nasa.gov

Cliff Neuman
Information Sciences Institute
bcn@isi.edu

Michael O'Connor
Brookhaven National Laboratory
moc@bnl.gov

Bob Oldham
NASA Goddard Space Flight Center
boldham@nastg.gsfc.nasa.gov

Mike Olsen
NASA Marshall Space Flight Center
mike.olsen@msfc.nasa.gov

George Pavel
Lawrence Livermore National Laboratory
gp@llnl.gov

Colin Perkins
Information Sciences Institute
csp@isi.edu

Dave Pleva
NASA Glenn Research Center
dpleva@grc.nasa.gov

Robert Poe
Business 2,0
rpoe@business2.com

Andrew Pohorille
NASA Ames Research Center
pohorill@raphael.arc.nasa.gov

Calton Pu
Georgia Institute of Technology
calton.pu@cc.gatech.edu

Mark Radwin
NASA Ames Research Center
mradwin@arc.nasa.gov

Mike Rechtenbaugh
U.S. Geological Survey
rech@edcmail.cr.usgs.gov

Russell Richards
National Oceanic and Atmospheric Administration/
Pacific Marine Environmental Laboratory
richards@pmel.noaa.gov

David Richardson
University of Washington
drr@washington.edu

Ronn Ritke
National Laboratory for Applied Network Research
ritke@nlanr.net

Matthew Rogge
Stanford University
mrogge@stanford.edu

Alain Roy
Argonne National Laboratory
roy@mcs.anl.gov

Michael Savaides
Qwest Communications
michael.savaides@qwest.com

Attendees cont.

Cathy Schulbach
NASA Ames Research Center
cschulbach@nas.nasa.gov

Mary Anne Scott
Department of Energy
scott@er.doe.gov

William Semancik
National Security Agency
wjseman@lts.ncsc.mil

George Seweryniak
Department of Energy
seweryni@er.doe.gov

Willard Smith
Tennessee State University
smith@coe.tsuniv.edu

Warren Smith
NASA Ames Research Center
wwsmith@nas.nasa.gov

Jerry Sobieski
Mid-Atlantic CrossRoads
jerrys@maxgigapop.net

Matt Chew
Spence NASA Ames Research Center
matt@nren.nasa.gov

Thom Stone
NASA Ames Research Center
tstone@arc.nasa.gov

Leigh Ann Tanner
NASA Ames Research Center
tanner@nas.nasa.gov

Ben Teitelbaum
Internet2
ben@internet2.edu

Jerry Toung
NASA Ames Research Center
jtoung@mail.arc.nasa.gov

William Turnbull
National Oceanic and Atmospheric Administration
wturnbull@hpcc.noaa.gov

Raymond Turney
NASA Ames Research Center
turney@nas.nasa.gov

George Uhl
NASA Goddard Space Flight Center
uhl.mamba.gsfc.nasa.gov

William Van Dalsem
NASA Ames Research Center
wvandalsem@mail.arc.nasa.gov

Michael Whisenant
CSOC Outsource
michael.whisenant@csoconline.com

Bessie Whitaker
NASA Marshall Space Flight Center
bessie.whitaker@msfc.nasa.gov

John Wroclawski
MIT
jtw@lcs.mit.edu

Glossary

2MASS	Two Micron Sky Survey
ALPHA	Advanced Laboratory for Parallel and High Performance Applications (JPL)
ANL	Argonne National Laboratory
AOX	All Optical Exchange
API	Application Programming Interface
ARC	Ames Research Center
ATDnet	Advanced Technology Demonstration Network
ATM	Asynchronous Transfer Mode
BNL	Brookhaven National Laboratory
CERN	European Organization for Nuclear Research, Geneva, Switzerland
CNRI	Corporation for National Research Initiatives
CPU	Central Processing Unit
DARPA	Defense Advanced Research Projects Agency
DHCP	Dynamic Host Configuration Protocol
DNS	Domain Name Service
DOE	Department of Energy
DPOSS	Digitized Palomar Observatory Sky Survey
DREN	Defense Research and Engineering Network
DS	Differentiated Services
DSTP	Data Space Transfer Protocol
DSVO	Digital Sky Virtual Observatory
DWDM	Dense Wavelength Division Multiplex
ECN	Explicit Congestion Notification
EGA	Early Gigabit Adopter
ELAN	Emulated LAN
ELT	Electronic Light Table
EOS	Earth Observing System
EPA	Environmental Protection Agency
FAQ	Frequently Asked Questions
FEC	Forward Error Correction
FPGA	Field Programmable Gate Array
Gbps	Gigabits per second
Gigaflop	One billion Floating Point Operations per second
GigaPOP	Gigabit Point of Presence
GN	Gigabit Networking
GPS	Global Positioning System
GSFC	Goddard Space Flight Center
HDTV	High Definition TV
HPC	High Performance Computing
HPCC	High Performance Computing and Communications
HPNAT	High Performance Network Applications Team (LSN)
HSCC	High Speed Connectivity Consortium
I2	Internet2
IP	Internet Protocol
IPMP	IP Measurement Protocol
IRAS	Infrared Astronomy Satellite
IS	Integrated Services
ISI	Information Sciences Institute (USC)
ISP	Internet Service Provider
JAMM	Java Agents for Monitoring and Management
JET	Joint Engineering Team (LSN)
JPL	Jet Propulsion Laboratory
KU	Kansas University
LAN	Local Area Network
LBNL	Lawrence Berkeley National Laboratory
LDAP	Light Directory Access Protocol
LLNL	Lawrence Livermore National Laboratory
LSN	Large Scale Networking
MAN	Metropolitan Area Network
Mbps	Megabits per second

Glossary cont.

MEMS	Micro-electromechanical Systems
MI	Middleware Initiative (Internet2)
MIB	Measurement Information Base
MONET	Metropolitan Optical Network
MPLS	Multi-Path Label Swapping
MTU	Maximum Transmission Unit
NAP	Network Access Point
NAS	Numerical Aerospace Simulation
NASA	National Aeronautics and Space Agency
NAT	Network Address Translation
NCO	National Coordination Office
NGI	Next Generation Internet
NGIX	Next Generation Internet Exchange
NIH	National Institutes of Health
NIST	National Institute of Standards and Technology
NLANR	National Laboratory for Applied Network Research
NLM	National Library of Medicine
NOAA	National Oceanic and Atmospheric Administration
NREN	NASA Research and Education Network
NRT	Networking Research Team (LSN)
NSF	National Science Foundation
NTON	National Transparent Optical Network
OC-12	Optical Carrier-12 (622 Mbps)
OC-3	Optical Carrier-3 (155 Mbps)
OC-48	Optical Carrier-48 (2.5 Gbps)
OCRL	Optical Communications Research Lab (Stanford)
OS	Operating System
PITAC	President's Information Technology Advisory Committee
PKI	Public Key Infrastructure
POC	Point of Contact
POS	Packet Over SONET
PPML	Predictive Model Markup Language
PVC	Permanent Virtual Circuit
QoS	Quality of Service
R&E	Research and Education
RFC	Request for Comments
RHIC	Relativistic Heavy Ion Collider
RSVP	Resource Reservation Protocol
SACK	Selective ACK (Acknowledgement)
SDSC	San Diego Supercomputing Center
SLA	Service Level Agreement
SNMP	Simple Network Management Protocol
SONET	Synchronous Optical Network
SoX	Southern Crossroads GigaPOP
SURA	Southeastern Universities Research Association
TCP	Transport Control Protocol
THz	Terahertz
UBR	Unspecified Bit Rate
UDP	User Datagram Protocol
USC	University of Southern California
UW	University of Washington
vBNS	Very High Performance Backbone Network Service
VBR	Variable Bit Rate
VH	Visible Human
VIDE	V Integrated Development Environment
VLAN	Virtual LAN
VMS	Virtual Mechanosynthesis
VRVS	Virtual Room Videoconferencing System
WAN	Wide Area Network
WDM	Wavelength Division Multiplexing